Defence Research and Development Canada | Recherche et développement pour la défense Canada

DEFENCE R&D DÉFENSE

Multimodal Interfaces

Literature Review of Ecological Interface Design, Multimodal Perception and Attention, and Intelligent Adaptive Multimodal Interfaces

Wayne Giang
Sathya Santhakumaran
Ehsan Masnavi
Doug Glussich
Julianne Kline,
Fiona Chui
Catherine Burns
Jonathan Histon
John Zelek

Contract Project Manager: Dr. Catherine Burns
PWGSC Contract Number: w7711-098148/001/TOR Call-up: 8148-01

Contract Scientific Authority:
Robert Arrabito
(416) 635-2033

Defence R&D Canada
Contract Report
DRDC Toronto CR 2010-051
May 2010

Canadä

Multimodal Interfaces

Literature Review of Ecological Interface Design, Multimodal Perception and Attention, and Intelligent Adaptive Multimodal Interfaces

Wayne Giang, Sathya Santhakumaran, Ehsan Masnavi, Doug Glussich, Julianne Kline, Fiona Chui, Catherine Burns, Jonathan Histon, John Zelek

University of Waterloo

Prepared By:
University of Waterloo
Advanced Interface Design Laboratory, E2-1303N
200 University Avenue West
Waterloo, Ontario
Canada N2L 3G1
Contract Project Manager: Dr. Catherine Burns
PWGSC Contract Number: w7711-098148/001/TOR Call-up: 8148-01
CSA: Robert Arrabito, Defence Scientist

Defence R&D Canada – Toronto

Contract Report

DRDC Toronto CR 2010-051

May 2010

Principal Author

Original signed by Catherine Burns

Catherine Burns

Professor, P.Eng., Ph.D.

Approved by

Original signed by G. Robert Arrabito

G. Robert Arrabito

Contact Scientific Authority

Approved for release by

Original signed by Marko Jovanovic

Marko Jovanovic

for Chair, Knowledge and Information Management Committee

This work falls under partner Group 13QH (Command)

Abstract

To improve operational effectiveness for the Canadian Forces (CF), the Joint Unmanned Aerial Vehicle Surveillance Target Acquisition System (JUSTAS) project is acquiring a medium-altitude, long-endurance (MALE) uninhabited aerial vehicle (UAV). In support of the JUSTAS project, Defence Research and Development Canada (DRDC) – Toronto is investigating the human factors issues of UAV ground control stations (GCS) interfaces for UAVs and exploring possible solutions using multimodal displays. This report analyzes current literature on multimodal perception and psychology in the context of developing a GCS simulator to evaluate the efficacy of multimodal displays for controlling UAVs. The report discusses the application of Ecological Interface Design (EID) to multimodal interface development, multimodal information presentation in non-visual modalities, and issues and implications of using multiple sensory modalities (e.g. crossmodal effects). In addition, the role of Intelligent Adaptive Interfaces (IAI) with respect to multimodal interfaces and current problems with automation in commercial aircraft are addressed. Recommendations are provided to develop a program of research to enhance the design of GCS interfaces to support future requirements of the JUSTAS project.

Résumé

En vue d'améliorer l'efficacité opérationnelle des Forces canadiennes (FC), l’acquisition d’un engin télépiloté (UAV) moyenne altitude et longue endurance (MALE) est un des volets du projet Système interarmées de surveillance et d'acquisition d'objectifs au moyen de véhicules aériens sans pilote (JUSTAS). À l’appui du projet JUSTAS, Recherche et développement pour la défense Canada (RDDC) — Toronto effectue des recherches sur les problèmes relatifs aux facteurs humains des interfaces UAV pour les postes de contrôle au sol (PCS) d'UAV et sur les solutions possibles au moyen d’affichages multimodaux. Le présent rapport porte sur l'analyse de littérature existante sur la perception et la psychologie multimodales dans le cadre du développement d'un simulateur PCS en vue d'évaluer l'efficacité d'affichages multimodaux pour commander les UAV. Le rapport comporte également un examen de l'application de la conception d'interfaces écologiques (EID) au développement d'interfaces multimodales, de la présentation d'information multimodale dans des modes non visuels et de problèmes et répercussions relatifs à l'utilisation de modes sensoriels multiples (p. ex. effets intermodaux). Le rôle d'interfaces adaptatives intelligentes par rapport aux interfaces multimodales et les problèmes actuels avec l'automatisation à bord des aéronefs commerciaux sont également abordés. De plus, des suggestions relatives à la mise au point d'un programme de recherches visant à améliorer la conception des interfaces PCS à l’appui des exigences futures du projet JUSTAS sont faites.

This page intentionally left blank.

Executive summary

Multimodal Interfaces: Literature Review of Ecological Interface Design, Multimodal Perception and Attention, and Intelligent Adaptive Multimodal Interfaces

Wayne Giang; Sathya Santhakumaran; Ehsan Masnavi, Doug Glussich, Julianne Kline, Fiona Chui, Catherine Burns, Jonathan Histon, John Zelek; DRDC Toronto CR 2010-051; Defence R&D Canada – Toronto; May 2010.

Background: Uninhabited aerial vehicles (UAVs) are remotely controlled aircraft used for a variety of civilian and military applications including command, control, communications, computers, intelligence, surveillance and reconnaissance (C4ISR). To improve C4ISR capability, the Canadian Forces (CF) is acquiring a medium-altitude, long-endurance (MALE) UAV under the Joint Unmanned Aerial Vehicle Surveillance Target Acquisition System (JUSTAS) project. In support of the JUSTAS project, Defence Research and Development Canada (DRDC) – Toronto is investigating human factors issues of ground control station (GCS) interfaces for UAVs and exploring possible solutions to enhance operator performance using multimodal displays. This report reviews literature on multimodal perception and psychology in the context of designing and evaluating the efficacy of a multimodal GCS simulator for controlling UAVs.

Results: Different human factors issues arise with various methods of operating and controlling UAVs. UAVs that are manually flown (e.g., manual takeoffs and landings) from remote locations suffer from decreased operator performance due to loss of sensory cues valuable for flight control, delays in UAV control inherent in the data link, and difficulty in scanning the visual environment surrounding the UAV. In contrast, for UAVs that are highly automated (e.g., automated takeoff, landings and preprogrammed flight), the human factors issues are primarily related to issues with supervisory control such as problems in monitoring, decision making, and situation awareness.

Many of these human factors issues can benefit from multimodal displays (i.e., an interface that communicates through visual, auditory, and tactile senses). A multimodal interface can enhance sensory cues relative to traditional visual GCS interfaces. Multimodal displays can also support supervisory control by presenting complementary and redundant information through multiple sensory channels. At times, it is also advantageous to substitute visual presentation of information with auditory or tactile displays. Multimodal presentation of information is also effective at capturing attention and improving response times to events (Sarter, 2006).

While the research on multimodal displays appears promising, the mapping of information in non-visual modalities and the crossmodal effects when combining multiple modalities is not well understood. There already exists a large body of literature on visual and auditory interfaces, but much less research has been conducted on tactile displays. This report provides a review of tactile perception, tactile displays and selected areas of auditory perception. In addition, we

discuss the literature on the crossmodal effects between visual, auditory and tactile modalities. Models of attention control and orientation are also described, and a comparison of how different modalities can be used to communicate the urgency of a message is discussed.

Another gap in multimodal research is the discussion of systematic methods to produce multimodal displays. Currently, there is little guidance for designers on how to develop an effective multimodal display. One possible method for designing multimodal displays is the use of Ecological Interface Design (EID). EID's main function is to assist the operator with understanding the system's underlying constraints so that the operator is able to respond to abnormal events. This is done by mapping constraints onto perceptual objects. Currently, there has been very little work done on extending EID to non-visual interfaces. This report discusses the few instances when EID has been applied to auditory and tactile interface designs.

Finally, possible future extensions to multimodal interfaces are discussed through a review of Intelligent Adaptive Interfaces (IAI). These interfaces allow for the system to respond intelligently to the user's goals, adapting to better support the tasks that the user is attempting to accomplish. A review of current IAI systems is used to provide insight into how these systems can be used in conjunction with future multimodal interfaces to better support users.

Significance: This report provides a comprehensive review of several topics relevant to the development of multimodal displays including a review of tactile perception, a discussion on the design of multimodal displays using EID, and how multimodal displays can be used in conjunction with IAI in future applications. This report will serve as a foundational and introductory document for anyone interested in developing future multimodal interfaces for enhancing operator performance.

Future plans: The development of a program of research to enhance the design of GCS interfaces will be performed based on the recommendations of this literature review. In particular, this report will assist DRDC Toronto with the design and development of a study to evaluate the efficacy of multimodal interfaces relative to traditional visual interfaces in a UAV autoland scenario. The results of the study will provide recommendations to support future requirements of the JUSTAS project.

Sommaire

Multimodal Interfaces: Literature Review of Ecological Interface Design, Multimodal Perception and Attention, and Intelligent Adaptive Multimodal Interfaces

Wayne Giang; Sathya Santhakumaran; Ehsan Masnavi, Doug Glussich, Julianne Kline, Fiona Chui, Catherine Burns, Jonathan Histon, John Zelek; DRDC Toronto CR 2010-051; R & D pour la défense Canada – Toronto; Mai 2010.

Introduction ou contexte: Les engins télépilotés (UAV) sont des aéronefs commandés à distance qui servent à diverses applications civiles et militaires, dont le C4ISR (commandement, contrôle, communications, informatique, information, surveillance et reconnaissance). En vue d'améliorer leur capacité C4ISR, les Forces canadiennes font l'acquisition d'un UAV moyenne altitude et longue endurance dans le cadre du projet Système interarmées de surveillance et d'acquisition d'objectifs au moyen de véhicules aériens sans pilote (JUSTAS). À l'appui du projet JUSTAS, Recherche et développement pour la défense Canada (RDDC) - Toronto effectue des recherches sur les problèmes relatifs aux facteurs humains des interfaces UAV pour les postes de contrôle au sol (PCS) d'UAV et sur les solutions possibles pour améliorer le rendement de l'opérateur au moyen d'affichages multimodaux. Le présent rapport porte sur l'examen de la littérature existante sur la perception et la psychologie multimodales dans le cadre du développement d'un simulateur PCS multimodal pour commander les UAV et de l'évaluation de son efficacité.

Résultats: Divers facteurs humains entrent en jeu selon les diverses méthodes d'utilisation et de commande des UAV. La commande manuelle (p. ex. atterrissages et décollages manuels) des UAV à partir d'emplacements éloignés donne lieu à une diminution du rendement de l'opérateur en raison de la perte de points de repère importants pour le pilotage, des retards de la commande d'un UAV inhérents à la liaison des données et de la difficulté à scruter l'environnement visuel en périphérie du UAV. Par contre, en ce qui a trait aux UAV commandés de façon presque entièrement automatique (p. ex. atterrissage et décollage automatisés, vol préprogrammé), les problèmes relatifs aux facteurs humains se limitent aux problèmes de supervision, comme la surveillance, la prise de décision et la connaissance de la situation.

Bien de ces facteurs humains peuvent tirés avantage des affichages multimodaux (c.-à-d. une interface qui communique au moyen de la vue, de l'ouïe et du toucher). Une interface multimodale peut améliorer les points de repère comparativement aux interfaces PCS visuelles traditionnelles. Les affichages multimodaux peuvent aussi prendre en charge la supervision grâce à la présentation d'information complémentaire et redondante via plusieurs canaux sensoriels. Il peut également être avantageux de remplacer la présentation visuelle de l'information par des affichages auditifs ou tactiles. La présentation multimodale de l'information est également efficace pour capter l'attention de l'opérateur et améliorer le temps de réponse aux événements (Sarter, 2006).

Quoique les recherches relatives aux affichages multimodaux se montrent prometteuses, la mise en correspondance de l'information et les effets intermodaux lorsque de multiples modes sont combinés ne sont pas très bien compris. Il y a déjà beaucoup de documentation sur les interfaces visuelles et les interfaces auditives, mais beaucoup moins sur l'affichage tactile. Le présent rapport comporte un examen de la perception tactile, de l'affichage tactile et de domaines choisis de la perception auditive ainsi que de la littérature relative aux effets intermodaux entre les modes visuel, auditif et tactile. Une description de modèles de contrôle et d'orientation de l'attention est également donnée, et une comparaison des façons dont les différents modes peuvent être exploités pour transmettre l'urgence d'un message est examinée.

L'examen de méthodes systématiques de génération d'affichages multimodaux est une autre lacune de la recherche. Il n'y a actuellement que peu de guides pour aider les concepteurs à développer un affichage multimodal efficace. La conception d'interfaces écologiques (EID) est une des méthodes possibles pour la conception d'affichages multimodaux. Une EID a pour fonction principale d'aider l'opérateur à comprendre les contraintes sous-jacentes du système, grâce à la mise en correspondance des contraintes avec des objets perceptuels, pour que l'opérateur puisse réagir en cas d'événements anormaux. Les travaux sur l'application des EID aux interfaces non visuelles sont actuellement peu nombreux. Le présent rapport comporte un examen des quelques occurences où une EID a été appliquée à des conceptions d'interfaces auditives et tactiles.

Enfin, des extensions futures possibles aux interfaces multimodales sont examinées au moyen d'une évaluation d'interfaces adaptatives intelligentes (IAI). Ces interfaces permettent au système de tenir compte des buts de l'utilisateur et de s'adapter pour mieux assister l'utilisateur dans les tâches qu'il tente d'accomplir. Un examen de systèmes IAI existants donne une idée de la façon dont ces systèmes peuvent être utilisés de pair avec les interfaces multimodales futures pour mieux assister les utilisateurs.

Importance: Le présent rapport constitue un examen exhaustif de plusieurs sujets pertinents au développement d'affichages multimodaux, dont une analyse de la perception tactile, un examen de la conception de d'affichages multimodaux faisant appel à l'EID et la façon dont les affichages multimodaux peuvent être utilisés de pair avec les IAI pour des applications futures. Il constitue une documentation de référence de base pour ceux intéressés à développer des interfaces multimodales futures en vue d'améliorer le rendement de l'opérateur.

Perspectives: L'élaboration d'un programme de recherches visant à améliorer la conception d'interfaces PCS se fera en fonction des recommandations faites dans le présent examen de la littérature. Notamment, le présent rapport aidera RDDC Toronto dans la conception et la mise au point d'une étude pour évaluer l'efficacité d'interfaces multimodales comparativement aux interfaces visuelles traditionnelles dans un scénario d'atterrissage automatique d'UAV. Les résultats de ces recherches apporteront des recommandations pour assister les exigences futures du projet JUSTAS.

Table of contents

List of figures

List of tables

This page intentionally left blank.

1 Introduction

Recently, there has been increased development and use of Uninhabited Aerial Vehicles (better known as UAVs) or Uninhabited Aerial Systems (UAS) to increase the capabilities of military and civilian forces in command, control, communications, computers, intelligence, surveillance, and reconnaissance (C4ISR) activities. These vehicle systems consist of remotely controlled/autonomous vehicles and Ground Control Stations (GCS) that provide C4ISR information without the need of carrying a pilot. This reduces risk, the need for on-board life support systems, weight, and fuel consumption, thereby increasing the range and possibilities of use of these vehicles. To improve C4ISR capability, the Canadian Forces (CF) is acquiring a medium-altitude, long-endurance (MALE) UAV under the Joint Unmanned Aerial Vehicle Surveillance Target Acquisition System (JUSTAS) project. In support of the JUSTAS project, Defence Research and Development Canada (DRDC) – Toronto is investigating human factors issues of GCS interfaces for UAVs and exploring possible solutions to enhance operator performance using multimodal displays. This report reviews literature on multimodal perception and psychology in the context of designing and evaluating the efficacy of a multimodal GCS simulator for controlling UAVs.

UAVs that are manually flown (e.g. manual takeoffs and landings) from remote locations suffer from decreased operator performance due to loss of sensory cues valuable for flight control, delays in UAV control inherent in the data link, and difficulty in scanning the visual environment surrounding the UAV. In contrast, for UAVs that are highly automated (e.g., automated takeoff, landings and preprogrammed flight), the human factors issues are primarily related to issues with supervisory control such as problems in monitoring, decision making, and situation awareness. Current UAS still require the use of an operator who is responsible for supervising and, if necessary, intervening in certain more critical situations such as during take-off and landing. All of these tasks may be supported through the use of multimodal displays, displays which communicate information through visual, auditory, and tactile senses. A multimodal interface can enhance sensory cues relative to traditional visual GCS interfaces. Multimodal displays can also support supervisory control by presenting complementary and redundant information through multiple sensory channels. At times, it is also advantageous to substitute visual presentation of information with auditory or tactile displays. Multimodal presentation of information is also effective at capturing attention and improving response times to events (Sarter, 2006).

The purpose of this project was to review the current literature in multimodal perception and displays to identify key findings on how to use multimodal technology effectively to improve UAV operator performance. This area encompassed such a broad range of topics that refinement of the literature occurred to support the specific objectives of the project as they became more clearly understood. The review of multimodal literature was focused on tactile perception, auditory display design, and crossmodal research as much of this research is quite recent and developing. Furthermore, the tactile literature was refined to focus on vibrotactile displays. The auditory research was focused to consider the research that was directly relevant to Ecological Interface Design (EID), crossmodal attention control, and auditory urgency and alarms. A set of specific research questions was developed (these are presented in Section 1.3) and refined the literature search even further.

In this introduction we discuss the literature review objectives, scope and structure. In the second section, we discuss EID, but limited very specifically to the multimodal applications of EID. As there are currently relatively few applications in this area, there is significant potential for EID to contribute in a meaningful way to the design of a multimodal GCS interface. In the third section we discuss the perception of vibrotactile displays. This is followed by a discussion of auditory display design in the fourth section. This section concludes with a discussion of how urgency information can be presented to operators across different sensory modalities. Section five describes potential perceptual issues that could occur when integrating tactile, auditory, and visual displays into a single multimodal display. In the sixth section of this report we review *Intelligent Adaptive Interface* (IAI) design, but restricted to the discussion of adaptive interface design in a multimodal context. Again, there is relatively little research in this area, reflecting the novelty of this application. Finally we conclude this report with recommendations for a program of research in multimodal interface design for UAV landing scenarios.

1.1 Literature Review Objectives

The objectives of this literature review in respect to the project's goals are as follows:

- To provide human factors advanced interface expertise for the design and development of a software-based GCS simulator in order to investigate the efficacy of multimodal displays for controlling UAVs.
- To perform a preliminary literature review of multimodal (auditory, visual, and tactile) perception and psychology.
- To perform a preliminary literature review of EID in multimodal displays and multimodal applications in general.
- To perform a preliminary literature review of adaptive interfaces, focused on adaptive multimodal display.

1.2 Literature Review Approach

Our preliminary literature review was initiated with a broad search of multimodal psychology and perception. We quickly found the need to refine this search to focus on questions that could be of interest within the following constraints:

1. We were interested in UAV automated landing situations.
2. We anticipated integrating the multimodal interface with an existing visual interface. This suggested that understanding cue conflicts and modality conflicts may be a promising direction to explore further.
3. We anticipate the multimodal interface to include a tactor vest providing vibrotactile signals and an auditory interface.

We used these constraints to further narrow the literature into specific questions of interest as follows:

1. What types of relationships can we show between variables in each modality? This includes subtopics like modality strengths/weaknesses and complementary/competing modalities along with current extensions to EID.

2. What occurs when two sources of information from different modalities conflict? Subtopics include visual dominance, crossmodal attention, short term conflicts vs. long term conflicts, etc.
3. What roles do adaptive interfaces play in the future of multimodal interfaces?

1.3 Multimodal Literature Review Structure

The structure of the report was developed in conjunction with the statement of work (SOW) and in relation to the established research questions established above. A copy of the statement of work is shown below:

- The items listed below will be executed by the Contractor:
    - Perform a preliminary literature review of multimodal (auditory, visual, and tactile) perception and psychology. Items to address include, but are not limited to:
        - Identify which modalities compete and which complement;
        - Identify most effective modality in the context of a GCS interface and information mappings;
        - Identify costs and confusions of modality switching;
        - Identify synaesthesia of modalities.
    - Perform a preliminary literature review of ecological interface design (EID) in multimodal displays and multimodal applications in general. Items to address include, but are not limited to:
        - Identify applications most relevant to this environment;
        - Determine whether EID can be used to derive insight into modality of information display;
        - Determine whether EID needs to be enhanced to generate these insights; how?
    - Perform a preliminary literature review of adaptive interfaces, focused on adaptive multimodal display. Items to address include, but are not limited to:
        - Determine whether there is any adaptive interface work that adapts the modality of display;
        - Indicate any adaptation rules that have been explored;

- Determine appropriate adaptation design guidelines to multimodal displays.

▪ Review and make recommendations for the baseline GCS interface that is being developed by DRDC Ottawa in order to ensure good human factors principles.

1.3.1 Categorization

Relevant literature was collected from scientific data bases, internal literature reviews, and scientific authorities. All articles were classified and evaluated in terms of type of paper, degree of peer review, modality and domain. The chart below depicts the tagging scheme used to organize the literature.

Type of Paper	Degree of Peer Review	Modality	Domain
- Event report - Technology review - Conceptual framework - Lab experiment - Simulator experiment - Field experiment - Literature review - Manual - Technical Report	- No peer review - Cursory peer-review - Intense, critical peer-review	- Visual - Auditory - Tactile - Multisensory - Spatial - Vestibular	- Military - Healthcare - Business - Energy systems - Transportation - Other

The degree of peer review was determined by examining the publication which the paper belonged to; conference proceedings and tech reports were assigned a cursory peer-review tag, journal articles and books were given an intense, critical peer-review tag, and all other articles were assigned a no peer review tag.

All articles were entered in a database called Mendeley which is a useful tool designed to store and organize literature and share them amongst group members. One of the many features included in Mendeley is the ability to highlight and make notes on PDF files, tag articles, and share them amongst group members. Access to the final Mendeley database will be provided to the Scientific Authority at the conclusion of this literature review.

1.4 Report Overview

This report is composed of six sections:

1. Ecological Interface Design and its applications to multimodal interface design:

   This section examines how the EID methodology is used in current visual interfaces, and how it has been adapted for use with non-visual interfaces. Future extensions to the methodology for use with multiple modalities are considered.

2. Tactile Perception:

   This section examines how individuals perceive information in the tactile modality, with a focus on vibrotactile stimuli. Tactile perception is described in detail due to the relative infancy of this branch of information presentation, and the lack of thorough discussion within the scientific community, in comparison to visual and auditory displays.

3. Auditory Display Design and Urgency:

   This section describes current research efforts directed towards auditory display design with a focus on how individual interpret auditory stimuli. A comparison of different methods of auditory information coding is provided, with a comparison with similar coding methods in the visual and tactile modalities. This section also provides a comparison of how urgency information is presented in different modalities.

4. Crossmodal Attention:

   This section describes how attention is directed in cases where an individual is presented with stimuli in multiple modalities. Different models of crossmodal attention are discussed, as well as issues dealing with interactions between stimuli in different modalities.

5. Intelligent Adaptive Interfaces (IAI):

   This section describes current research in the realm of IAIs. It also draws on the research described in the previous sections of the literature review to envision how multimodal interfaces can also be adapted to better support the user's goals.

6. Developing a program of research:

   This final section describes information that is directly relevant to the design of a study to evaluate the benefits of using a multimodal interface. This includes a cognitive walkthrough of autoland mission scenarios, and experimental methodologies that deal with similar multimodal/automation supervision tasks. Finally, a number of possible experiments are presented.

2 Ecological Interface Design

Ecological interface design (EID) is a design approach that has been used to great success in complex socio-technical systems (Vicente, 2002; Vicente & Rasmussen, 1992). In this section we provide a brief review of the EID design approach, its goals and principles, and we review the few cases where EID has been applied to non-visual interfaces. An extension of EID for auditory displays developed by Sanderson, Anderson, and Watson (2000) will be described in further detail, and its implications for multimodal design will be discussed. Finally, we examine some possible crossmodal issues which still need to be addressed by EID and consider how EID can be used to further benefit multimodal interface development.

This section is organized as follows:

- Section 2.1. Provides a background of the goals behind EID and how EID benefits interface design.
- Section 2.2. Describes current lines of research using the EID methodology that have been done with non-visual interfaces.
- Section 2.3. Describes the semantic mapping process.
- Section 2.4. Describes the *attentional mapping* process.
- Section 2.5. Extends the EID design process to include support non-visual modalities.
- Section 2.6. Provides insights into possible shortcomings in the EID methodology that need to be addressed in order to better support multimodal interface design.
- Section 2.7. Provides concluding remarks about EID.

2.1 Background

EID is a design methodology that is focused on supporting the control and monitoring of large systems by supporting an operator's understanding of the underlying constraints of the system (Vicente & Rasmussen, 1992). EID also focuses on ecological sound interfaces, which are "designed to reflect the constraints of the work environment in a way that is perceptually available to the people who use it." (Burns & Hajdukiewicz, 2000, p. 1) This differs from other design methodologies such as the 'user centered design' technique. These methodsare analyzed from the perspective of the user and are not as focused on the system as a whole. However, interface designers who use the EID methodology must have a complete understanding of the system, and this is done through the use of a number of analyses techniques such as a *Work Domain Analysis* (WDA) and *Control Task Analysis* (CTA), which are also used in the *Cognitive Work Analysis* (CWA) framework.

The EID framework is built on top of Rasmussen's *skills, rules, knowledge* (SRK) taxonomy which describes different levels of cognitive control. Operators of complex systems are capable of using control strategies based on *Skill-Based Behaviour* (SBB), *Rule-Based Behaviour* (RBB), or *Knowledge-Based Behaviour* (KBB). SBB represents behaviour that arises due to extensive training and experience, resulting in almost automatic responses to incoming signals. RBB occurs when operators are able to follow a rule or procedure. KBB exists when events that are unforeseen by both of the operator and designers occur, and operators must use their knowledge of the system to diagnosis the problem.

Vicente and Rasmussen (1992) describe three fundamental principles of design that make use of the SRK taxonomy:

- SBB: To support interaction via time-space signals, the operator should be able to act directly on the display and, the structure of the displayed information should be isomorphic to the part-whole structure of movements.
- RBB: Provide a consistent one-to-one mapping between the work domain constraints and the cues or signs provided by the interface.
- KBB: Represent the work domain in the form of an abstraction hierarchy to serve as an externalized mental model that will support knowledge-based problem solving

By supporting all three levels of cognitive control, operators are able to choose the lowest level of control required for the task at hand, while still allowing intuitive access to more detailed information when required. Vicente (2002) reviewed a number of interfaces which made use of the design methodology. He found that in practice, EID provided performance increases in terms of increased speed at resolving faults, and decreased variability in results. These were the result of a number of different factors. First and foremost is the re-organization of information using the abstraction hierarchy. This re-organization allowed the operator to control the level of complexity of the system by viewing it at different levels of abstraction. Vicente showed that organization of information using the abstraction hierarchy resulted in performance improvements even in the absence of different visual forms. Secondly, the unique visual forms that are used to support RBB improved performance by loading spatial processing rather than verbal processing. Vicente and Rasmussen (1992) also state that perceptual judgements have reduced variability when compared to analytical judgements. Thus, the benefits of using EID come from presenting required information at different levels of abstraction, and by using perceptual judgements to represent the constraints and relationships between the different levels.

In the past, EID has been used in a variety of different domains including transportation systems (command and control of a frigate, command and control of a destroyer, and displays for aircraft), process control systems (thermal power generation systems, nuclear power simulations, acetylene hydrogenation reactors), telecommunication systems (network management), and medical systems (oxygenation monitoring in the neonatal intensive care unit, patient monitoring in the operating room, diabetes management) (Burns & Hajdukiewicz, 2000). These application areas all involve complex systems with constraints that are often not known by individual users, and allowed the users to explore the data at different levels of complexity.

2.2 Multimodal Applications of EID

While the majority of research done using EID has been done using visual displays, the framework is not restricted only to the visual modality (Vicente, 2002). However, there have been relatively few researchers who have extended EID to other modalities. The following table provides a list of these lines of research.

Table 1: Lines of multimodal research using EID.

Papers	Application Domain	Modalities Used	Extensions to EID	End Results
Lee, Stoner, and Marshall (2004)	Driving	Haptic, Visual	Comparisons of driving scenarios to process-control scenarios	Guidelines for haptic design based on SRK
Davies, Burns, and Pinder (2007)	Sonar mobility devices	Auditory	Comparisons	Prototype interface (Usability study / Cognitive walkthrough evaluation)
Watson, Anderson, and Sanderson (2000)	Aircraft landing and approaches	Auditory, Visual	Attentional Mapping	Sonification for landing (not tested)
Sanderson, Anderson, and Watson (2000); Watson, Anderson, and Sanderson (2000); Sanderson, and Watson (2005); Watson and Sanderson (2007); Anderson and Sanderson (2009)	Anaesthesia	Auditory, Visual	Extended Design Process, Attentional Mapping	Sonification anaesthesia interface (non-clinical tests)

As can be seen, the majority of the non-visual research has been done in the auditory modality. None of the research done has resulted in testing the EID interface against interfaces designed using other design methodologies. In fact, the majority of the research has not been formally evaluated in published studies. The research done by Sanderson, Anderson and Watson is on-going, and consists of the most complete extension of the EID process to date. Out of the four domains of research that have been explored using non-visual EID interfaces, one of these (Davies et al., 2007) focuses on only the auditory modality. This was done because the project was modelled after sonar systems that have previously been designed for visually impaired individuals. The other three projects all consist of some degree of presentation in multiple modalities. This is because the application domains that were used (driving, anaesthesia, and to a lesser degree aircraft landings) are tasks which the operators gather a portion of the required information through direct haptic perception of the environment.

Lee et al. (2004) argue that this direct perception is an important difference between the driving scenario and the process control scenarios in which EID interfaces have been employed. In the process-control environment, operators rarely have a chance to directly interact with the components they are monitoring; their information is normally mediated through the interface. However, in driving the car, the operator is exposed to a number of multisensory cues that come directly from the environment. Therefore, they suggest that signals that have “direct analogical links to the signals from the driving environment” should be used to promote SBB. Similar results were found by Davies et al. (2007) when they found that *auditory icons*, sounds which have a direct link to a real world object or event (such as footsteps), performed better than *earcons*, sounds which do not have a direct link to the real world but can be arranged to communicate information. In the anaesthesia *sonification* designed by Watson and Sanderson (2007), the tempo of breath inspiration and expiration was used to help communicate data in a manner that took advantage of the fact that anaesthesiologists are already sensitive to the breathing patterns of their patients. **Taken together, these findings suggest that skill-based behaviour is easiest to support when the signal has some real-world link that the operators are already sensitive to.**

2.3 Semantic Mapping

Semantic mapping is a process where variables are mapped into perceptual characteristics. This process is fundamental to fulfilling the 2nd EID principle where constraints should be mapped onto perceptual objects (Vicente & Rasmussen, 1992). Since humans are fine tuned to detect certain perceptual changes, changes in conditions that take a system out of a safe area can trigger RBB if the changes also cause the perceptual object to become more salient. Sanderson et al. (2000, p. 62) describe the following list of seven heuristics by Hansen (1995):

1. Goal achievement as figural goodness.
2. Work domain constraints as visual containers.
3. Process dynamics as figural changes.
4. Functional relations as visual connections.
5. Pictorial symbols to represent components.
6. Alphanumerical output where needed.
7. Time as visual perspective.

Sanderson et al. (2000) adapted four of these heuristics into the auditory domain:

- *Goal achievement as figural goodness*: Sanderson et al. equated the concept of figural goodness in visual stimuli to acoustic simplicity.

- *Work domain constraints as visual containers*: Containers are a spatial concept that Sanderson et al. state is difficult to replicate in the auditory domain.

- *Process dynamics as figural changes*: In the auditory domain this could be represented by changes in acoustic parameters.

- *Functional relations as visual connections*: Relationship of different acoustic parameters to each other.

Using these concepts, Watson et al. (2000) developed an auditory sonification for aircraft landing and approach which is very applicable to the UAV GCS scenario. In the WDA for this scenario, two types of variables were categorized: those related to spatial location (altitude, air speed and direction), and those related to "engineering function" (control of thrust and automation). The auditory system is capable of doing spatial recognition as well as differentiating between different characteristics of the auditory stream. The authors reformatted the landing task into an auditory pursuit task, where the ideal glide slope was mapped into a spatial location around the operator. Thus, any non-central location would indicate that the plane has gone off its ideal glide slope, triggering RBB. The pursuit task is also an example of SBB because the operator is able to make adjustments based on the direction of the sonification. The engineering functions were mapped onto the auditory characteristics of the sonification as shown in Figure 1. Air speed was represented as the "time between iterations of all four engines (the tempo of the sound)". The engine setting, which indicates the direction of thrust, was represented using differences in pitch relative to a reference pitch. Other auditory characteristics such as reverberation and timbre were also mapped onto events which required attention (reverse thrust, and changes in settings).

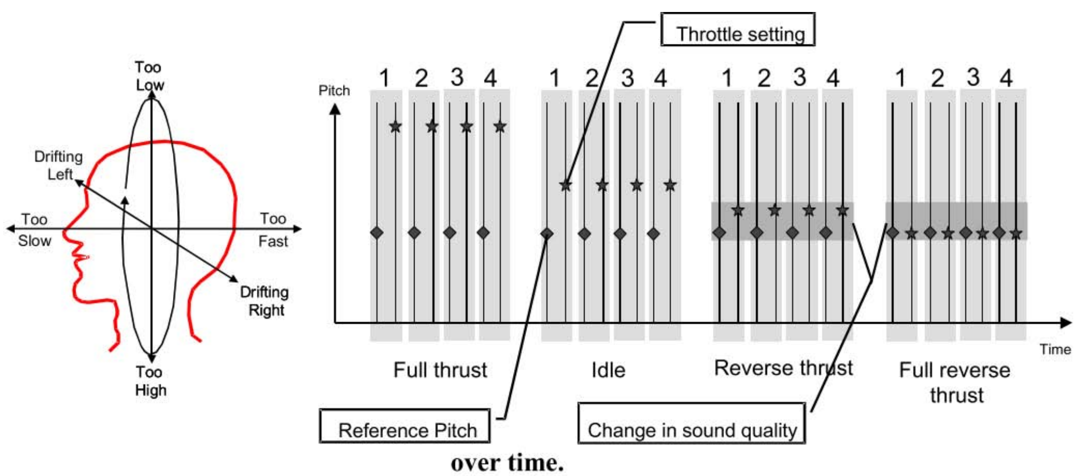


Figure 1: Representation of the aircraft approach and landing sonification. Taken from Watson, Sanderson, and Anderson (2000, p. 7)

The work done to map work domain variables to perceptual qualities is one that requires extensive knowledge of both the application domain and perceptual characteristics of the modality used. Burns and Hajdukiewicz (2004) describe the use of a *visual thesaurus* to assist with this difficult mapping problem. The visual thesaurus is a set of visual forms that can be used to represent work domain properties. The visual forms used include visual primitives (bar graphs and other simple iconic elements), complex combinations of visual primitives (connections, grouping, etc.). By using these individual elements a "visual ecology" can be created that allow the operator to process information about system constraints based on visual perceptual judgements.

Sanderson and Watson (2005) consider the concept of an auditory thesaurus built on the use of *earcons, auditory icons, audifications*, and *sonifications*. (An audification is a straight signal-to-sound conversion, whereas as a sonification is a mapping of information to sound parameters to create the auditory equivalent of a visualization). They also discuss the literature relevant to the next steps of the EID methodology suggested by Burns and Hajdukiewicz, which has been adapted in the following table.

Table 2: Methods for selecting correct auditory stimuli (Sanderson & Watson, 2005)

Step	Description	Psychoacoustic Literature
Range of variation and critical variables	Choose parameters and perceptual dimensions that are capable of showing the range of values required, and show context information such as critical values or boundaries.	How directions of measured values should be mapped onto auditory dimensions (Walker, 2002) Calibration of auditory dimensions for anaesthesia values (Anderson & Sanderson, 2004; Anderson & Sanderson, 2009)
Relationships between multiple variables	Show how individual variables are related to each other.	Auditory scene analysis (Bregman, 1990) Single stream or multiple streams? (Anderson & Sanderson, 2004)
Means-end relations	Display the links between the different levels in the abstraction hierarchy.	Auditory scene analysis (Bregman, 1990)

The majority of the work that has been completed so far has largely relied on fulfilling the first step of choosing parameters with the correct range for displaying information. The relationships between different variables are still not adequately represented. For example, in the landing scenario described above, the relationship between engine speed and thrust should relate to the position of the optimal glide slope, but this relationship is one that must be calculated by the operator. **Displaying relationships between data (either between variables or the means-end relations) is currently the largest challenge faced in mapping semantic information into non-visual modalities. It may be worthwhile to consider tactile displays in terms of *tactons*, tactile icons, direct signal to tactile mappings, and mapping of information to tactile parameters similar to a sonification or visualization.**

2.4 Attentional Mapping

One of the extensions proposed by Sanderson et al. (2000) was the inclusion of an attentional mapping step to EID. When working with a single modality, such as vision, a designer can largely assume that the operator's attention will be focused on the display when the information is required. However, as the number of channels of information increases, the assumption of focused attention on any one display is no longer valid. This is true even for purely visual displays that are spread out over many monitors, or if a task also requires observation of non-display elements in the environment. Visual displays also tend to be localized, optional, and persistent. Operators are able to refer to them when required, while ignoring them when they are not needed. However, auditory data possesses three characteristics, it is ubiquitous, obligatory, and transitory (Sanderson & Watson, 2005). While some tactile displays are not always in contact with the skin, those that are in constant contact with the skin also share the same characteristics as auditory displays. Unlike visual displays where inattention can lead to missed stimuli, tactile displays are ubiquitous. An operator has much less choice in whether they want to attend to these ubiquitous displays, and the salience of the display becomes much more important. It is important to note that tactile displays refer to displays constructed from vibrotactile tactors which are in contact with the skin. Tactile displays provide "passive" feedback, in contrast to haptic feedback which traditionally refers to force-feedback or other types of responses that occurs with "active" touch (where a user actively reaches out to interact with an object).

Sanderson et al. (2000) proposed that information about the *attentional profiles* of the operators need to be gathered as part of the analysis phase in EID. To this end, they recommended the use of other portions of CWA such as CTA, Strategies Analysis (StA), and Social Organization Analysis (SOA). The CTA, and its variant the temporal coordination control task analysis (TC-CTA) is especially relevant to building attentional profiles, which can be used to develop a sense of what data the operators should be focused on during different tasks. Sanderson et al. (2000) suggest that ubiquitous displays should work both while the display is in focal awareness, and when it is outside of focal awareness, as shown in Figure 2.

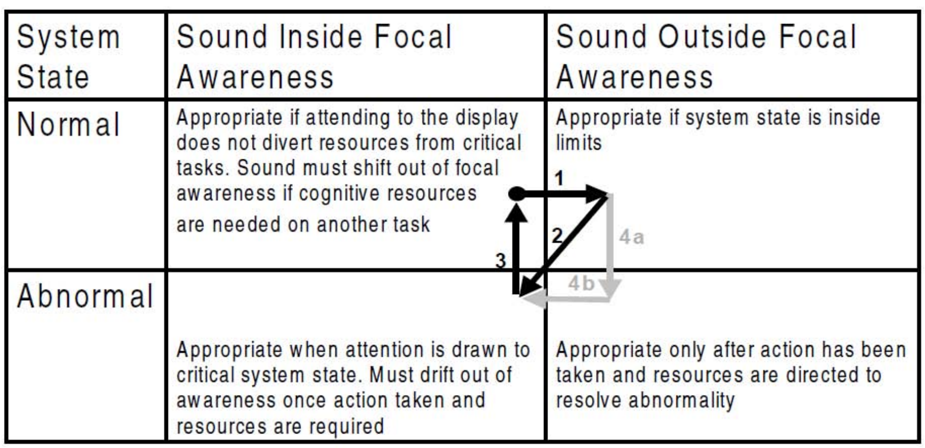


Figure 2: Exploiting Auditory Attention (modified from Sanderson et al., 2000, p. 265)

The attentional mapping helps determine when the display should be within focal attention, the boundaries where it should transition out of focal attention, and when it should attempt to capture the operator's attention to bring itself back to focal awareness. One of the advantages of using this method for directing attention is that it allows for smooth transitions in and out of focal attention. Alarms are highly salient and are designed to readily capture attention, but they are often distracting and can degrade performance during times of high cognitive load (Woods, 1995 as cited in Sanderson et al., 2000). **The consequences of this step are that the designer needs to have an understanding of which perceptual characteristics can be processed pre-attentively (so that they can be used outside of focal awareness), which properties can capture attention (so that the operator can orient their attention when required), and what characteristics can provide the required bandwidth of information transfer while in focal attention. The concept of temporal-coordination control task analysis may be useful in building this understanding.**

2.5 Design Process Extensions

As part of their extensions of EID for the design of auditory displays, Sanderson and Watson (2005; Watson and Sanderson, 2007) developed a design process that assists with gathering the requirements needed for an auditory display. A graphical representation of this process, taken from Watson and Sanderson (2007) can be seen in Figure 3.

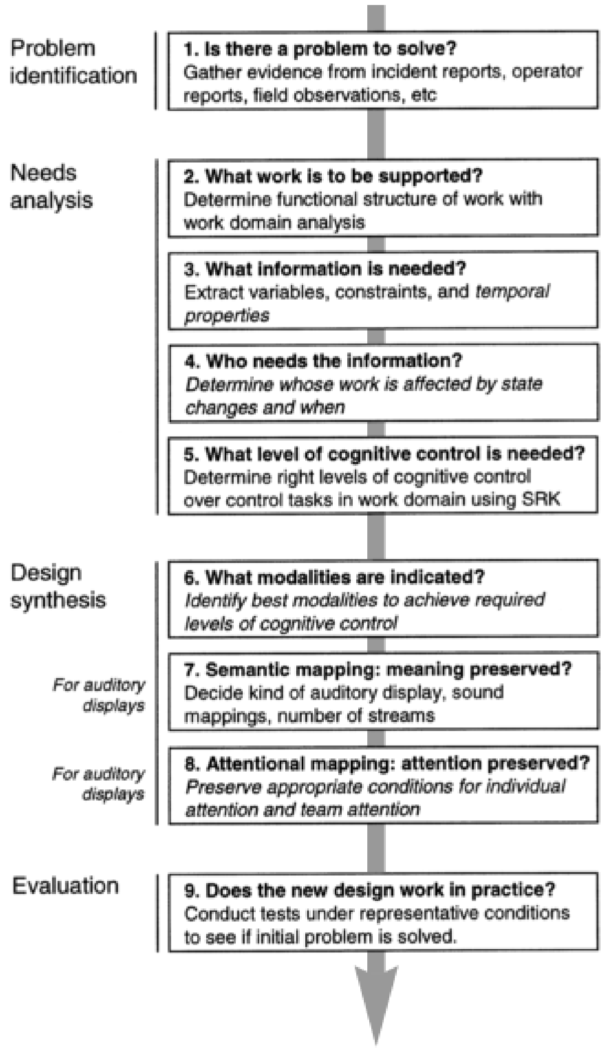


Figure 3: Auditory EID design process (Watson and Sanderson, 2007, p. 2)

This is a practical design process which can be followed in the design of any auditory display, and with some modifications to the semantic mapping step of the process, it could also be used for a tactile display. Since each sensory modality allows an operator to perceive different types of stimuli, the semantic mapping step is unique for each modality. While some parallels can be drawn between different modalities, such as with oscillatory signals in both the auditory and tactile modalities, interface designers must be careful not to assume that a semantic mapping in one modality works in another.

There are also a few areas where this design process could be extended to improve the requirements gathering process for a multimodal interface. The first possible extension is a mapping of the sensory stimuli that operators are already presented with. This would provide some insight into the ambient stimuli in the work environment, while also providing information about what types of sensory stimuli operators are already using. A report by Williams (2008) on the sensory information that is available or lacking in the operation of UAV systems may provide valuable insight into the types of sensory information that should be provided.

The second problem with the current design methodology is that it treats each modality relatively independently, while only checking for crossmodal interactions during the evaluation stage. This greatly simplifies the design process, since crossmodal interactions have largely only been studied in laboratory environments (see Sarter, 2006 for a review, as well as the Section 5 in this report). However, the concepts of semantic mapping, and attentional mapping could extend beyond a single modality. Some possible examples include the use of a multisensory cue to capture attention, or a variable that is mapped into an auditory dimension and a visual dimension. Whether these would lead to performance improvements over current “modality-independent” interfaces is still an open research question.

2.6 Crossmodal Implications for EID

Sarter (2006) reviews a number of current multimodal interface guidelines, and one of the major problems is that they do not address a number of crossmodal interaction problems:

- *Modality expectations*: If an operator expects a cue to appear in a certain modality, they experience “enhanced readiness to detect and discriminate information in that sensory channel.”
- *Modality shifting effect*: Operators have difficulty shifting their attention away from an expected modality to a modality that contains less frequent targets.
- *Crossmodal attention shifting*: Shifts in spatial attention in one modality also tend to shift attention in other modalities.
- *Exogenous and endogenous attention*: In real-world tasks, an operator will have goal-driven (endogenous) responses to stimuli, but the interface is also able to capture attention using stimuli-driven (exogenous) cues. The interaction between these two forms of attention is still not well understood.

The current extended EID methodology still does not have the tools to explicitly deal with these problems. However, many of these crossmodal issues can be added to the attentional mapping step to help guide the direction of focal attention so that these issues can be avoided. More importantly, a formal crossmodal interaction evaluation should be conducted at the end of the design process to ensure that information channels that are meant to be independent do not interact in a detrimental manner.

Finally, it is important to reconsider the foundations of EID to examine what elements provide the most benefit to operators. The current research has largely focused on designing perceptual forms in other modalities that designers can leverage to support RBB. However, the ability of auditory and tactile displays to show relationships between data in one modality is still very limited. Configural displays are one method of showing these relationships, but the concept of different levels of abstraction in different modality channels has not yet been explored in the literature. Burns (2000) explored how spatial and temporal proximity affects an operator's ability to integrate information in a visual display. She found that high spatial proximity provided the largest benefit in the operator's ability to diagnose faults in a process control task, and this was improved when the interface also had high temporal proximity. While spatial proximity is a concept that can also be applied across different modalities, auditory and tactile information may be less dependent on spatial information than visual information. A study of information integration across different modalities may provide insight into whether provision of different levels of abstraction through different modalities is a valid design option.

2.7 Concluding Remarks

In conclusion,

- The EID approach provides benefits both due to the re-organization of information using means-ends links, and because of changing analytical judgements into perceptual judgements.
- There is currently very little research done on extending EID to other modalities, and the multimodal interfaces that have been designed using this method have not been tested against multimodal interfaces designed using other methodologies.
- *Auditory signals* (earcons, auditory icons, audifications, and sonification) provide a ripe lexicon of perceptual signals that can be used by designers to support SBB, RBB, and KBB.
- Attentional mapping is a step that is important when designing modalities that cannot be ignored, and that have strong temporal qualities.
- EID designers must have a strong understanding of perceptual dimensions, and how these perceptual dimensions can direct the operator's attention.
- The current literature on configurable displays to support relationships between variables and means-ends links is still in its infancy.
- It may be possible to support the re-organization of information into different modalities to support different levels of abstraction. However, this would require displaying means-ends links across modalities. Further research into how operators integrate abstract information across modalities is required.

3 Tactile Perception

Drawing from some of the insights gained from the previous section on EID, it is clear that a strong understanding of how operators perceive information in different modalities is important for interface designers. Vision and audition are the two best understood modalities which humans use to interact with the outside world. These modalities can provide highly precise spatial and temporal information. Thus, the field of human-computer interface design and human factors engineering has focused much of their study and design on these modalities (Lederman & Klatzky, 2009; Van Veen & Van Erp, 2000). On the other hand, the sense of touch has been largely ignored despite the fact that it is an essential part of human ability to interact with the environment. We are particularly interested in developing a strong foundation of tactile perception research because it can assist with the effective use of the vibrotactile vest. For this reason, we have started at an anatomical level. Subtle effects of tactor stimulation of the skin, such as adaptation rates and discrimination and localization ability, can have implications on how tactile displays should be designed. The review demonstrated that key findings from the basic science governing the sense of touch are relevant to interface design. This section also includes guidelines regarding vibrotactile parameters which can be used in generating tactile messages using the vest. It is important to note that other types of tactile and haptic interfaces exist (Bliss, Katcher, Rogers, & Shepard, 1970; Priplata, Niemi, Harry, Lipsitz, & Collins, 2003; Galvin, Mavrias, Moore, Cowan, Blamey, & Clark, 1999), but they are beyond the scope of this literature review.

This section is organized as follows:

- Section 3.1. Provides an anatomical overview of human skin to provide insight into how tactors produce sensation. This understanding becomes important for effective *tacton* design.

- Section 3.2. Discusses the effects of vibrotactile stimuli placement, and localization issues on the torso. These effects are important in understanding how to design tactile signals in a tactor vest.

- Section 3.3. Describes vibrotactile spatial acuity of the trunk and the effects of vibrotactile timing parameters on localization performance. This provides the foundation for the basic design of tactile signals.

- Section 3.4. Provides guidelines for coding information through vibrotactile displays.

- Section 3.5. Discusses different types of vibrotactile patterns along with a discussion of the research results.

- Section 3.6. Reviews other tactile characteristics.

- Section 3.7. Discusses the current understanding of masking effects in tactile displays.

- Section 3.8. Presents concluding remarks and summary of tactile perception.

3.1 Anatomic Overview of the Skin

We know from experience that a simple tap can immediately draw our attention. The nervous system is very capable of spatially localizing stimuli on the skin. For this reason, stimulation of the skin can be a powerful way to passively convey spatial information. The surface of the body might play an important role in presenting information to operators in situations where their other senses are being used or overloaded (Van Veen & Van Erp, 2000). In the last few years, there has been rapid growing interest in the development and application of interfaces which use tactile technology as a way of communicating spatial and navigational information to operators (Rupert, 2000; Van Veen & Van Erp, 2000; Van Erp, Van Veen, Jansen & Dobbins, 2005).

The anatomical characteristics of human skin receptors have been discussed in detail in numerous reviews (Kandel, Schwartz & Jessel, 1991; Greenspan & Bolanowski, 1996; Cholewiak & Collins, 1991). Only a brief summary is provided in this section in order to provide a basic understanding of how tactile displays influence the body. Skin is the largest receptive organ on the human body (Chouvardas, Miliou & Hatalis, 2005). There are various receptor structures buried deep in the multi-layered tissue of the skin. In order to design applicable interfaces, the understanding of the various sensitivities of the skin's sensors and their responses to external stimuli is helpful. To date, the majority of studies of tactile interfaces have focused on mechano-receptors located within the glabrous (hairless) skin of the human. As Figure 4 depicts, underneath the surface of the glabrous skin, three thin layers exist: The first layer is the epidermis and its thickness varies from 0.4 mm to 1.6 mm. The second layer is the dermis which is about 6 times thicker than the epidermis and the third one is the subcutis (hypodermis)(Lederman & Klatzky, 2009; Chouvardas et al., 2005).

The skin contains a variety of sensory organs called receptors. These are divided into 5 main groups by the type of stimuli that they are sensitive to: *mechanoreceptors* which are sensitive to pressure, vibration and slip, *thermoreceptors* which are sensitive to changes in temperature, *nocioreceptors* which are pain receptors, and *proprioceptors* which give information about the position of the limb in space. Various receptors respond to particular vibration frequencies and have different tendencies to adapt to vibratory stimuli. Frequency and adaptation characteristics should be considered in the design of tactile displays.

Referring to Figure 4, four kinds of mechanoreceptors lie in the skin tissue, each at specific depths of the skin (Cheung, Van Erp, and Cholewiak, 2008; Sherrick & Cholewiak, 1986; Lederman & Klatzky, 2009):

- *Meissner corpuscles* are a stack of nerve fibres, located in the grooved projections of the skin surface formed by epidermal ridges, situated perpendicular to the skin surface. They respond to light touch and are velocity sensitive. They are sensitive to vibrotactile stimuli in the range of 10 – 100Hz. They have highest sensitivity (lowest threshold) when sensing vibrations less than 50Hz. Meissner corpuscles are categorized as *rapid adapting* (RA) receptors which respond quickly to a stimulus, but rapidly adapt to it and stop responding when subjected to a constant stimulus.

- *Merkel receptors* are disk shaped receptors that respond to pressure and texture, but also to low frequency (5-15 Hz) vibratory input. They are categorized as *slow adapting* (SA) receptors which adapt slowly to stimulus and continue to transmit when subjected to constant pressure. Tactile display systems, by necessity, are in constant contact with the skin and are not well suited for the stimulation of SA type receptors.

- *Ruffini corpuscles* are spindle shaped receptors that respond to skin stretch and mechanical deformation within joints, specifically angle changes up to 2 degrees. They contribute to providing feedback for the grip and grasping function. These are categorized as SA receptors and are located in the deep layers of the skin.

- The *Pacinian corpuscles* are the largest receptors of the skin. These are located deeper in the skin and most susceptible to the vibrations in the 200-350 Hz frequency range. Pacinian corpuscles are categorized as RA receptors. This means that the effect of stimuli degrades rapidly after onset. Pacinian corpuscles discharge only once per stimulus application, hence they are not sensitive to steady pressure.

In general, the most effective and applicable receptors in tactile display applications are the Merkel cells for pressure sensation, the Meisner corpuscle for low frequency and the Pacinian corpuscle for high-frequency vibrations (Chouvardas et al., 2005). The most relevant receptors for the design of the tactile vest, which make use of C2 tactors operating at an optimal frequency of 250 Hz, are Pacinian corpuscles.

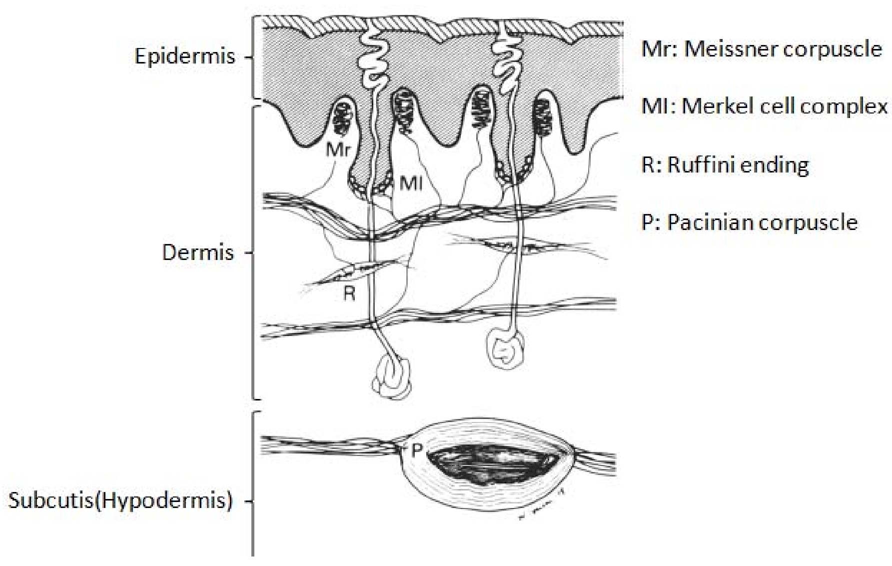


Figure 4: Glabrous skin anatomy. Picture taken from Lederman and Klatzky (2009, p. 1440).

3.2 The Effects of Placement on Vibrotactile Localization on the Torso

There have been several attempts since the 19th century to investigate the spatial acuity of the skin on several body parts. Generally, as we move from distal regions (such as the hands) to proximal regions (such as the torso) of the body, the sensitivity to stimuli degrades. The law of mobility states that the skin's sensitivity to locating and discriminating touched locations improves as the mobility of parts of the body increase (Cholewiak, Brill, & Schwab, 2004; Van Erp, 2005b). In addition to this, vibratory stimuli can be localized more effectively when they are located on anatomical points of reference. For example, when Cholewiak and Collins (2003) evaluated vibratory stimuli localization at the various sites of the arm, they concluded that stimuli were localized best when they were presented near the wrist, elbow, and shoulder. **As a result, when developing tactile displays where spatial localization should be optimized, the design should consider taking advantage of anatomical points of reference to improve localization.**

The majority of research that has attempted to investigate the accuracy and limitations of the sense of touch has typically tended to present stimuli to more sensitive regions of skin, such as hands and finger tips (Cholewiak & Collins, 2003; Hillstrom, Shapiro, & Spence, 2002). Although hands may have better discriminative power than the rest of the body, most current interfaces already require the use of the operator's hands and limbs for control activities. This fact highlights the importance of investigating the potential for using the surface of the torso as an alternative way to convey information. The three-dimensional nature of the body presents a natural mapping for three-dimensional spatial information (Gallace, Tan, & Spence, 2007). Individuals tend to use the orientation of the trunk as a frame of reference in determining their self-orientation. This is because the head and limbs do not provide a stable frame of reference because they rotate relative to the trunk (Karnath, Schenkel, & Fischer, 1991). Therefore, knowing the effects of space and place on the vibrotactile localization on the torso is essential.

Several comprehensive experiments have been performed by Van Erp as well as Cholewiak and his colleagues. These researchers have investigated the ability for individuals to localize vibratory stimuli around the torso (Cholewiak et al., 2004; Van Erp, 2005a). Both of these experiments were conducted with the use of vibrotactors. Vibrations are commonly used as stimuli since the skin rapidly adapts to stationary touch and pressure (Nafe & Wagoner, 1941). "Adaptation may be generally defined as a reduction in sensitivity resulting from a continuous unchanging stimulus" (Cheung, Van Erp, & Cholewiak, 2008, p. 2-4).Therefore, taps on the skin have to be repeated in order to create a vibratory stimulus that the skin will not adapt to. In general, people can distinguish a temporal gap of 5 ms between successive taps on the skin (Lederman & Klatzky, 2009). Pressure based stimuli is more susceptible to adaptation, and is only sensitive to Merkel receptors. Vibrotactile stimuli on the other hand can be sensed by Pacinian corpuscles, the largest of the receptor structures in the skin. It is important to note that Cholewiak et al. used the same C2 tactors which have been used in the tactile vest of our project.

3.2.1 Torso Location and Localization

Arrays of vibrotactors can be used to represent the location of an object relative to body in the environment. Cholewiak et al. (2004), in the first part of their experiment, presented stimuli using vibrotactors situated at 12 equidistant locations on two belts. The belts encircled the abdomen and the lower margin of the rib. The reason for using two levels (abdomen and lower margin of the rib) was to see whether the characteristics of the underlying tissue would affect the localization of the vibrotactile stimuli. The vibrotactors located on the frontal side of the lower belt was placed on the tissue of the abdomen, whereas vibrotactors of the upper belt were over the ribs. In each trial, one stimulus (vibrotactor) was activated.

The first portion of the experiment revealed that the participant's performance in detecting stimuli around the abdomen and the rib cage was similar. Therefore for the torso, the underlying tissue type plays a minor role in vibrotactile spatial location. The ability to localize a stimulus around the torso was found to be a function of proximity to the spine (6 o'clock) and the navel (12 o'clock). It was found that observers were more capable of correctly detecting stimulus near the spine (6 o'clock) and the navel (12 o'clock) and these points can serve as anatomical reference points for the trunk. **For this reason, in designing tactile displays, the spine and the navel could be used as reference locations in spatial tactile displays.**

3.2.2 Tactor Separation and Localization

In the second part of the Cholewiak et al. (2004) experiment, the number of vibrotactors on the belt was varied to evaluate whether better localization performance is possible with a decreased number of tactors. This was inspired by information transmitted and channel capacity of the observer notions described by Miller (1956). Arrays of 8 and 6 tactors were used as test conditions. The results of the second part of the experiment, compared against those obtained with the 12-tactor condition in the first part are shown in the polar plot presented in Figure 5. Overall performance around the torso was found to be dramatically improved when the number of vibrotactors was reduced, though there was still variation in performance based on the location of the tactor.

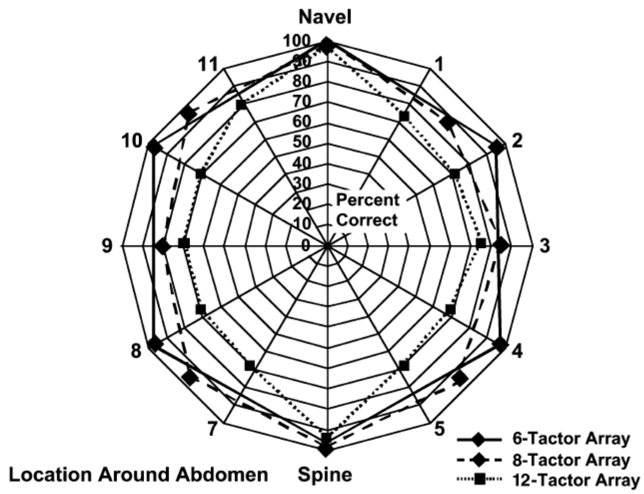


Figure 5: Localization performance around the abdomen for 6, 8, and 12 vibrotactile belts. Figure taken from Cholewiak et al. (2004, p. 979).

Similarly to the first part of the experiment, participants had the highest level of performance when localizing stimuli which were located at the navel and the spine when compared to other locations on the torso. This was true for the 6, 8, and 12 tactor array conditions. The results of this experiment suggest that increasing the separation between tactors and thus decreasing the number of vibratory stimuli improves the localization performance dramatically. **In consideration of this, tactile pattern designs should take into consideration that increased tactor separation and reduced stimuli may improve localization performance.**

In order to demonstrate the importance of the spine and navel anchor points as points of reference, the vibrotactors belt arrays were rotated slightly so that tactors fell on the sides of these points. As shown in Figure 6, in both these cases, the performance decreased.

In the third part of the experiment, 7 vibrotactors were located on a short strip spanning roughly half the circumference of the body and this tactor strip was used in 4 locations on the torso: front, back, left side and right side of the body. In the first case the array across the abdomen (front) was arranged so tactor 1 was at the left, tactor 4 at the navel and tactor 7 at the right side. For the back case, tactor 1 was at the right side, tactor 4 at the spine and tactor 7 at the left side of the body. The other two cases had similar orientations, but had tactors that started at the navel or spine, and a center tactor (tactor 4) on either the left or right side of the body. The results of these experiments are depicted in Figure 7. Better performance was obtained when the tactor strip was used on the front and back, when compared to when it was located on the left side or right side of the body.

Summarizing the results of the Cholewiak et al. (2004) experiments we can derive three main conclusions:

1. **The spine and the navel can work as natural anchor points and observers are more capable of correctly detecting stimulus near these points.**

2. **Performance is found to be dependent on the number of tactors around the body, therefore increasing the separation among the tactors improves the localization ability.**

3. **Individuals are better able to localize tactors placed on the front and back of the torso than either the left or right sides of the body.**

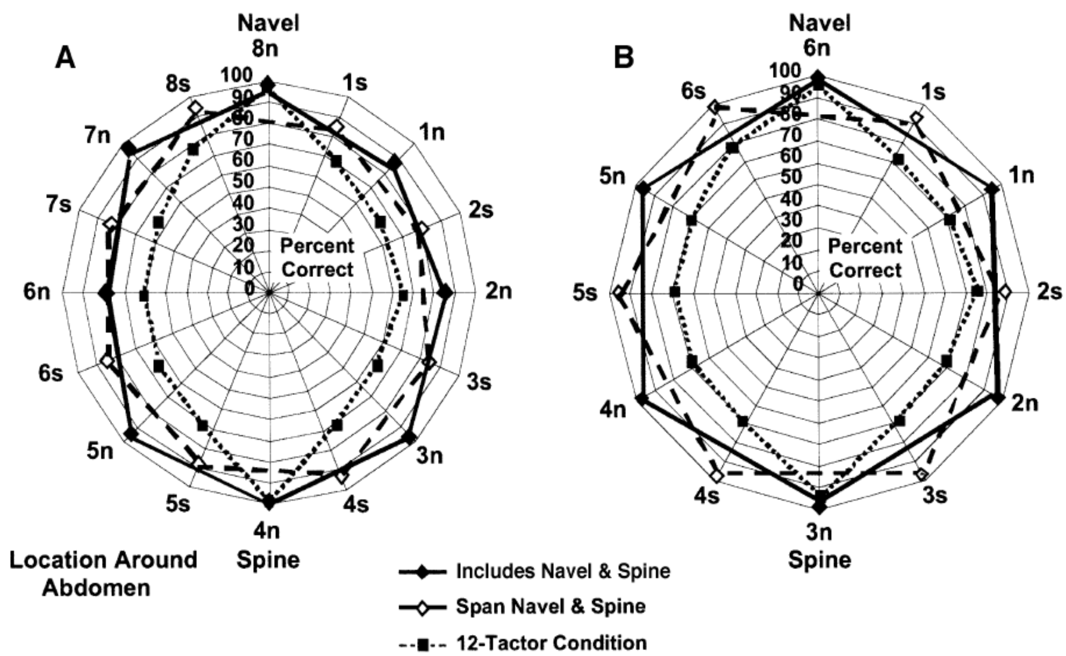


Figure 6: Localization performance around the abdomen; A) for 8 tactors and B) for 6 tactors. The solid lines in each graph connect the performances for the conditions that two of the tactors were situated on the spine and the navel (n); dashed lines connect the performances for the condition that the navel and the spine were spanned (s). The data represented by the dotted lines are from the first part of the experiment (12 tactor condition). Figures taken from Cholewiak et al. (2004, p. 980).

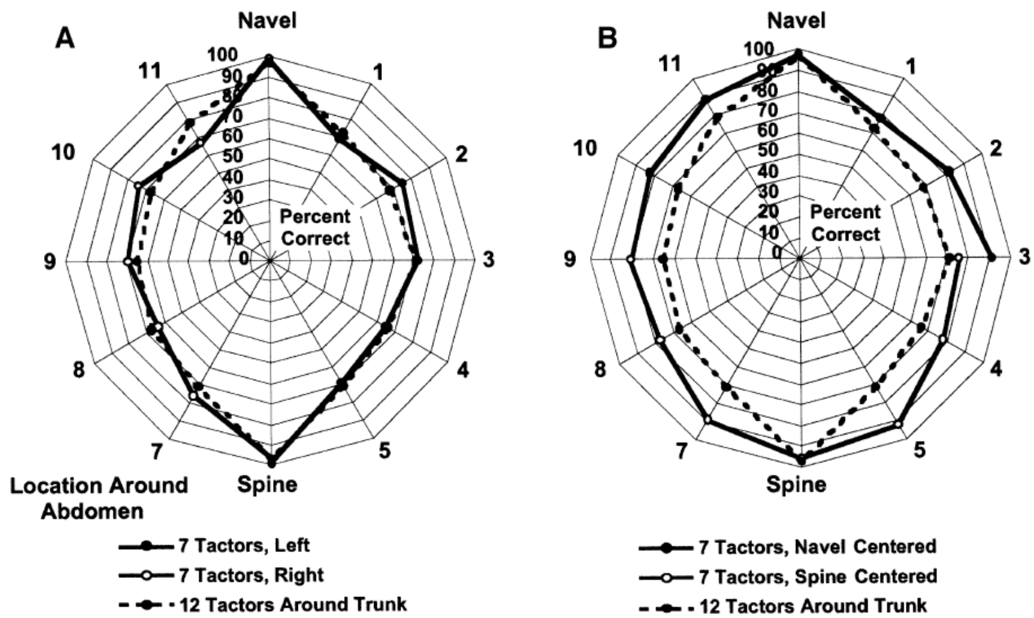


Figure 7: Localization performance for seven tactors presented to seven sites of the body in 4 cases. Figures taken from Cholewiak et al. (2004, p. 983).

3.2.3 Origin of Reference Points for Tactor Localization

In another study by Van Erp (2005a), participants wore a tactor belt consisting of 15 vibrotactors. Tactors were embedded equidistantly around the belt's circumference. The middle tactor was located just above the navel. One stimulus, consisting of a vibrating tactor, was activated in each trial. The participants were asked to indicate the location of the vibration on a horizontally positioned square board, which they were seated within (by means of a specialized apparatus which was designed for this experiment). Figure 8 shows the results of this experiment. Van Erp (2005a) found that there was a bias between the actual location of the tactors on the torso and the locations indicated by the participants as their response. The bias was toward the midsagittal plane, that is, perceived locations were located towards the navel for the tactors located on the abdomen and towards the spine for the tactors located on the back. This result is consistent with the findings of Cholewiak et al. (2004) and supports the fact that the navel and the spine can be considered as anchor points of the torso.

All participants showed a pattern in which the lines from the indicated location of the tactor on the square board to the actual tactor spot on the observer's body surface seemed to cross at one of two points. One of these points exists for the left and one for the right half of the body, with a mean lateral distance of 6.0 cm between them.

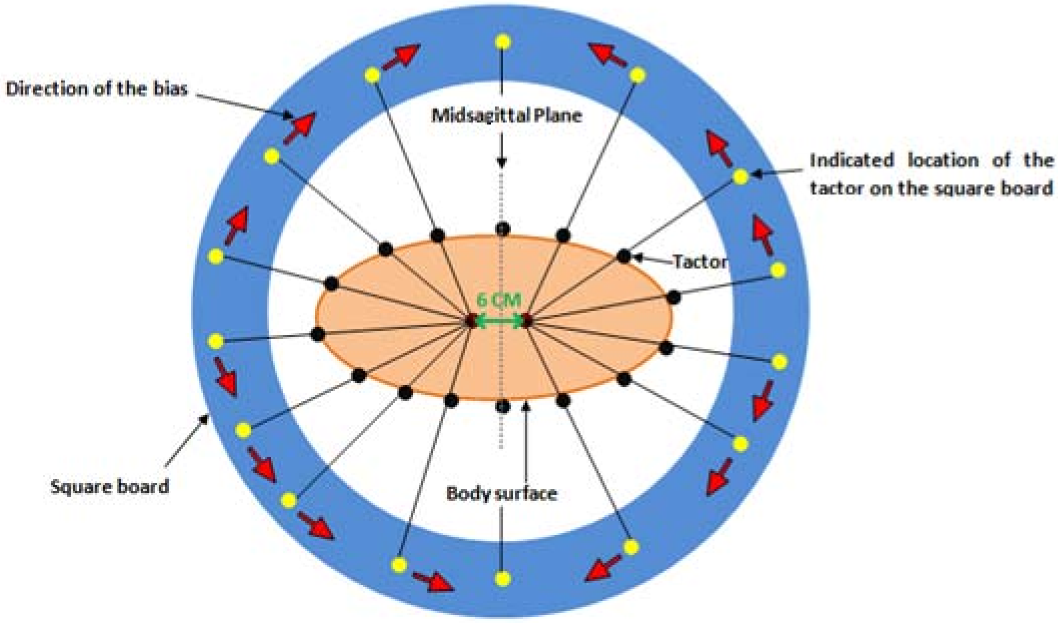


Figure 8: Schematic top view of the Van Erp's experiment and results. Red flashes indicate the direction of the bias in the response of participants. Adapted from Van Erp (2005a, p. 307).

Summarizing the findings of the Van Erp (2005a) experiment resulted in two main conclusions:

1. **The navel and spine can be considered anchor points of the torso.**

2. **There are two internal reference points in the human body, one for each half (left and right), and observers do not use the center of the torso as the origin for observed direction. This suggests that spatial tactile signals should be designed from the internal reference points in the body, and not simply from the midsagittal plane as this reflects how people tend to interpret the signals.**

3.3 Vibrotactile Spatial Acuity of the Trunk and Effects of Timing Parameters on Localization Performance

Spatial acuity has been investigated by several methods and most studies have used pressure or brief touches instead of vibrotactile stimuli (Cholewiak & Collins, 2003). Weinstein measured thresholds of *two-point discrimination* (minimum distance between two stimuli to be perceived as two distinct stimuli instead of one large stimulus) and tactile point localization on several body locations using pressure stimuli (Weinstein, 1968). The lowest thresholds were found for the finger tips and were found to be 2.5 mm for two point discrimination and 1.5 mm for point localization. In contrast, thresholds for the trunk were larger and were found to be around 4 cm for the back and 3.5 cm for the abdomen for two-point discrimination and 10 mm for point localization (for both the back and abdomen). Pressure stimuli are detected by Merkel receptors, but vibrotactile stimuli are detected by Pacinian corpuscles which results in different spatial acuities for the two different types of stimuli. Considering our project uses vibrotactile stimuli on a vest, it would be pertinent to include the results of the Van Erp investigations about the acuity of the torso in discrimination of vibrotactile stimuli which will be presented in the following sections (Van Erp, 2005b).

3.3.1 Spatial Acuity by Location

In the first part of the Van Erp's experiments the spatial resolution of vibrotactile stimuli on different locations of the torso was investigated (Van Erp, 2005b). This was done by placing vertical and horizontal arrays of tactors on the skin of the back and abdomen. In this experiment, each presentation consisted of the sequential activation of two vibrotactors. The experimental task was to indicate whether the second tactor was presented to the left or to the right of the first tactor for the horizontal arrays, and above or below of the first tactor for the vertical arrays.

The results of this experiment demonstrated a uniform acuity of about 2-3 cm across the trunk and there were no acuity differences between horizontally and vertically located arrays. These values are similar to the findings of Weinstein who found spatial acuity of the trunk to be around 3-4 cm for pressure stimuli (Weinstein, 1968). The acuity was better for horizontally oriented arrays located on the spine and the navel and was about 1 cm for these regions. This midline accuracy provides further evidence that the spine and the navel can serve as anatomical anchor points as was demonstrated previously by Cholewiak et al. (2004), not just because they are anatomical reference points, but because acuity may also be more accurate in these locations. **For the design of tactile signals, active tactors should be at least 3 cm apart on the torso, and 1 cm apart on the navel or spine. The navel and spine regions may provide better acuity, reinforcing the idea that these areas may serve as good reference points.**

3.3.2 Spatial Acuity and Timing

In the second part of the Van Erp (2005b) experiment, the effects of the timing parameters on localization performance were assessed. Before we continue, we need to define two concepts:

- *Burst Duration (BD):* which is the time between the onset and end of a burst

- *Stimulus Onset Asynchrony (SOA):* which is the time between the onsets of two consecutive bursts

Four pairs of tactors were attached to the back of participants as can be seen in Figure 9. The center-to-center distance between two tactors within a pair was 2.5 cm. The distance between two pairs was 3.5 cm. Each presentation consisted of the sequential activation of two tactors with 25 combinations of *BDs* and *SOAs*. The task of the observers remained the same; participants were asked to indicate whether the second tactor was to the left or to the right of the first tactor. The final results are depicted in Figure 10. Both *BD* and *SOA* were found to affect the localization performance of participants. Performance improved when *BD* and *SOA* increased, and *SOA* was found to have larger effects on performance than *BD*. Therefore, there is a trade-off between the speed of stimulus presentation and spatial acuity. Hence, **applications which utilize tactile displays and require high spatial acuity can profit from longer *BDs* and *SOAs,* and tasks that depend on fast response times should make use of larger distances between the vibrotactors (Van Erp, 2005b)**.

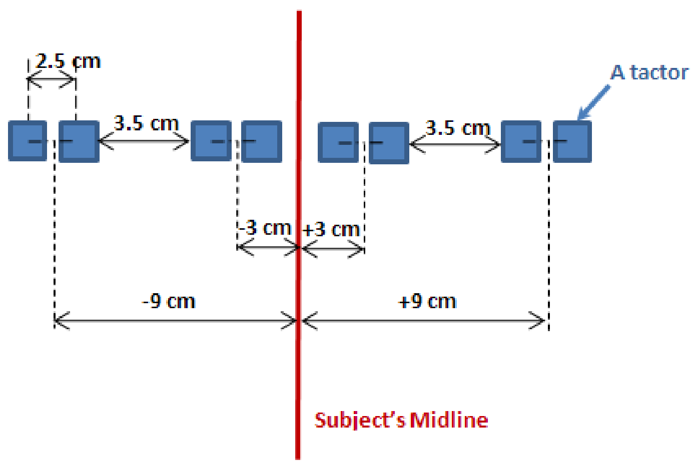


Figure 9: Placement of Tactors for Van Erp (2005b) Experiment

Summarizing the results of the Van Erp (2005b) experiment, we can derive two main conclusions:

1. **Spatial acuity is relatively uniform over the trunk and it is approximately 2-3 cm for vibrotactile stimuli. This acuity is better for horizontally oriented arrays located on the spine and navel and is about 1 cm for these regions.**

2. **Localization performance improves when *BD* and *SOA* of two sequentially activated vibrations increase.**

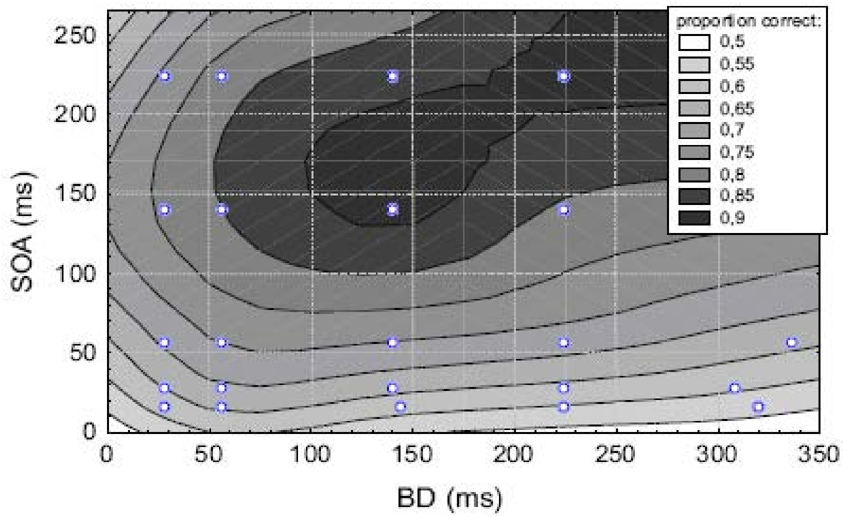


Figure 10: Effects of the timing parameters on localization performance. Proportion correct as function of BD and SOA. Darker colors indicate better performance. Figure taken from Van Erp (2005b, p. 83).

3.4 Guidelines for Coding Information through Vibrotactile Displays

The sense of touch is a unique communication channel and vibrotactile displays transfer information by presenting vibrations through this channel. The interest in application of vibrotactile displays is growing, and these displays have already been used in a number of applications:

- As a sensory substitution for people with visual or hearing disabilities. For example, Optacon is a device that translates written text into vibrotactile signals through an array of pins in contact with the user's finger (Bliss et al., 1970; Priplata et al., 2003; Galvin et al., 1999);
- To assist with orientation and navigational tasks for operators in situations where disorientation occurs due to mismatched vestibulo-ocular response and the absence of stable frames of reference (Van Veen & Van Erp, 2000; Van Erp & Van Veen, 2003; Van Erp & Van Veen, 2004; Van Erp, Van Veen, Jansen & Dobbins, 2005);
- As directional cues for areas of interest (Oskarsson et al., 2008);
- To help show the amount of deviation from a planned course, and to alert the operator to unexpected events (Donmez et al., 2008);
- For exploring computer-generated virtual environments;
- As omni-directional alerts and alarms (Calhoun et al., 2003; Calhoun et al., 2004).

Considering the many possible applications of vibrotactile displays, an investigation of different methods of information representation and coding principles (how to develop tactile patterns that can be understood within a specific application) would be pertinent. The focus of the following subsections is on how different tactile parameters can be manipulated to present messages in vibrotactile displays.

3.4.1 Coding Information by using Different Frequencies

Optimal sensitivity of human skin to vibration is within 150 to 300 Hz (Jones & Sarter, 2008). For frequencies outside of this interval, the displacement of the skin must be greater to be detected. The amplitude required for detecting vibration at any given frequency varies for different locations on the body. Wilska (1954) measured detection thresholds of 25-1280 Hz vibrations for different locations on the body. He found the lowest threshold amplitudes within the frequency range 200-450 Hz. For 200 Hz vibrations, the finger tips have the lowest threshold of 0.07 µm, whereas in the abdominal and gluteal regions the lowest detection threshold is as high as 14 µm (Sherrick & Cholewiak, 1986)**.**

Verrillo (1962; 1963) measured the sensitivity to vibration on the glabrous skin of the hand as a function of frequency, tactor properties, and differences in the pressure upon the skin. **Based on the results, the detection threshold as a function of frequency was found to be a U-shaped curve which has its minimum in the region of 250Hz.** He also demonstrated that threshold decreases as the vibrating contactor, the portion of the tactor in contact with the skin, pressed further into the skin. In another experiment, Verillo concluded that the size of the area of stimulation is a significant parameter of a vibrotactile stimulus. When the area was reduced, higher thresholds of detection were recorded (Verrillo, 1966). Cholewiak et al. (2004) measured vibrotactile detection thresholds as a function of stimulus frequency by presenting stimuli on 6 equidistant locations on a vibrotactile belt which encircled the abdomen. They reported that there is no statistically significant difference between vibrotactile detection thresholds around the trunk. A vibrotactile stimulus at a given frequency was perceived similarly at spine, navel and four additional loci on the sides of the abdomen. Figure 11 shows the results of this experiment. **Taken together, these results suggest that tactors do not require additional compensation or tuning to achieve similar levels of perceived vibration when used in the tactor vest.**

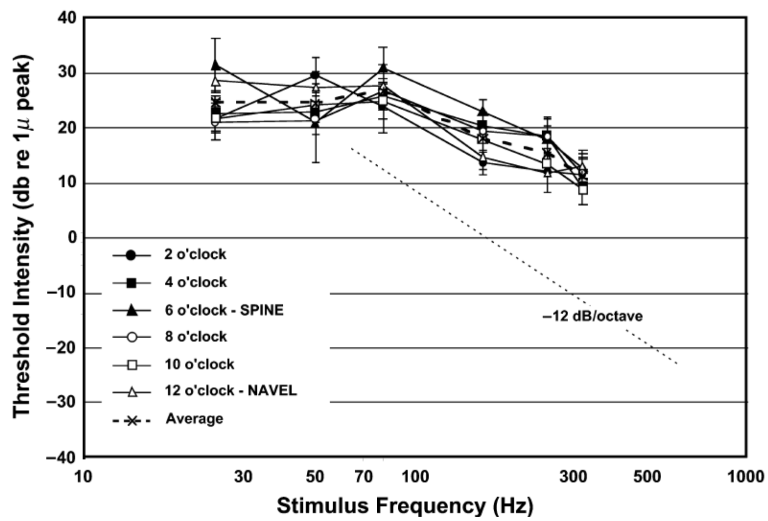


Figure 11: Vibrotactile detection thresholds measured at six locations around the abdomen. Figure taken from Cholewiak et al. (2004, p. 973).

There are no extensive studies on the ability for individuals to discriminate between different frequency levels of vibrotactile and we have relatively little data on this topic. Therefore it is difficult to specify distinct changes in vibrotactile frequency that could be correctly distinguished by operators (Jones & Sarter, 2008). Rothenberg, Verrillo, Zahorian, Brachman, and Bolanowski (1977) suggested that an appropriate scale of vibration frequency may include approximately seven differentiable levels from the lowest to the highest applicable values on the forearm. Sherrick (1985) presented vibrations to the finger and reported that within the frequency range of 2-300 Hz, between three to five levels of vibrotactile frequency can be discriminated by humans, and this can be increased up to eight recognizable levels when intensity is added as a redundant cue. He also found that discrimination above 100 Hz deteriorates rapidly. The results of this study also state that a low frequency vibration at high intensity can be incorrectly perceived as a moderate vibration at medium intensity. This highlights the fact that increasing the amplitude of a vibration also increases the perceived frequency of the signal (Jones & Sarter, 2008).

Other studies have suggested that a maximum of nine different levels of frequency should be used for coding information (Van Erp, 2002; Brewster & Brown, 2004). Also, differences between frequency levels for vibrations with equal amplitude should be at least 20% (Van Erp, 2002). Brewster and Brown (2004) also state that “the number of frequency steps that can be discriminated also depends on whether the vibrotactile cues are presented in a relative or absolute way. Making relative comparisons between stimuli is much easier than absolute identification, and this will lead to much fewer discriminable values.” It should be noted that for areas with less sensitivity and lower density of innervations like the trunk, increases in the perceived frequency grow more rapidly with increases in frequency of the physical stimuli (Jones & Sarter, 2008).

The Weber Fraction is a formula that is often used to determine the minimum threshold of perceived change in any parameter (e.g., amplitude, frequency, weight). For frequency, it is the differential threshold divided by the reference frequency, expressed as a percentage.

$$k = \frac{\Delta I}{I} \qquad (1)$$

Where *K* is the Weber Fraction, *I* is the reference amount of the parameter and *ΔI* is the minimum threshold of the perceived change in a parameter (e.g. frequency)

The Weber fraction is reported to change as a function of frequency. However, different results are reported from different authors. In one study the Weber fraction increased from about 18% at low frequencies to 30% at 300 Hz, whereas in another study it decreased from 30% at low frequencies to 13% at 200 Hz (Jones & Sarter, 2008).

Summers et al. (1997) investigated the perception of step changes in stimulus frequency. The stimuli were periodic signals of 80, 160, 240, and 320 ms durations with one octave step change of frequency at their halfway point. For example a signal of 240 ms duration was increased/decreased one octave in its frequency after 120 ms from its onset. There were also constant stimuli with no step change. Three different waveform types were used for this experiment: sine wave, monophasic pulse, and a tetra phasic pulse. Figure 12 illustrates these waveforms. Vibrations were presented at two different sensation levels, 24 dBSL and 36 dBSL. The experiment showed that participants were able to correctly detect constant stimuli, but with increasing or decreasing frequency of the stimuli there were more unsuccessful discriminations as shown in Figure 13.

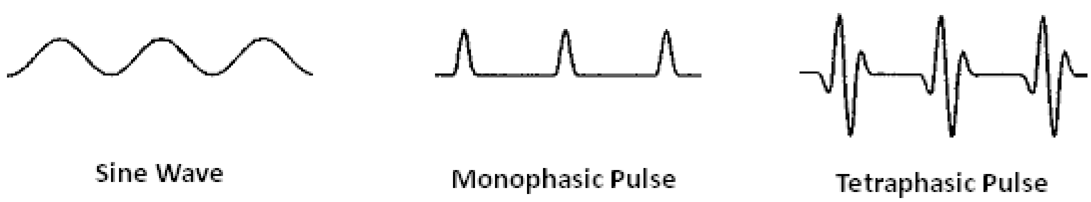


Figure 12: Three types of waveforms used in the Summers et al. (1997) experiment. Adapted from Summers et al. (1997, p. 3687)

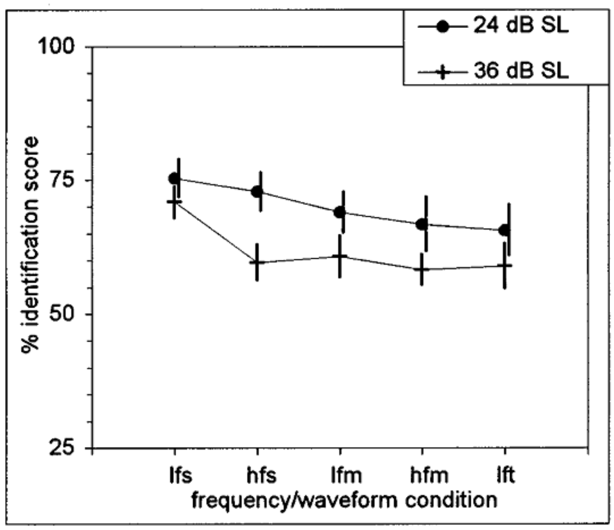


Figure 13: Overall results of the Summer et al. (1997) experiment. lfs = 50/100Hz sine; hfs = 200/400Hz sine; lfm = 50/100Hz monophasic; hfm = 20/400Hz monophasic; lft = 50/100Hz tetraphasic. Figure is taken from Summers et al. (1997, p. 3690).

Due to the large amount of variation and uncertainty about the perception of changes in frequency, changes in frequency may not be a useful method for presenting messages in vibrotactile displays. Also, the limited bandwidth of frequency of electrical devices and tactors may limit the display when information is coded using different frequency levels. Therefore frequency should be cautiously changed in these displays, especially when amplitude is also being manipulated as a variable (Jones & Sarter, 2008).

3.4.2 Coding Information by using Different Amplitudes

Changes in amplitude of vibration can be a very useful parameter to encode information in vibrotactile displays (Brown & Brewster, 2006a). For example, the urgency of a message can be represented by presenting vibrations with different amplitudes to the operator’s skin. Therefore, it is important to know how individuals are able to perceive different amplitudes of vibrations in terms of intensity or magnitude. One of the units of measurement for amplitude is *decibels above sensation level* (dBSL). It measures the amplitude of a signal relative to an individual’s sensation threshold. For example, if a person's minimum sensation threshold is 20 dB and a signal is at 40 dB, the sensation level of this signal for this individual is 20 dBSL. Craig (1972) measured the difference threshold (the minimum change in amplitude that can be discriminated by an individual 50% of the time) of a 160 Hz vibration presented to the right index finger. The signal was raised to 14, 21, 28, and 35 dBSL. He found that the difference threshold at these levels is constant and is approximately 1.5 dB. Craig (1972) also found that the difference threshold increases with decreasing intensity below 15 dBSL.

Verrillo et al. (1969) measured contours of equal-sensation of magnitude judgments, resulting from the interaction of frequency and amplitude. The stimuli consisted of 10 different vibrotactile frequencies and were presented by a 2.9 cm^2 contactor to the thenar eminence (i.e. palm) of the right hand. The experiment consisted of two main sections. In the first section, a series of 10 stimuli (one for each of 10 different vibration frequencies) with different amplitudes were randomly presented. Participants were instructed to assign numbers to each presented stimulus (magnitude estimation). In the second section, participants could control the amplitude of vibrations by means of a control knob. They were instructed to adjust the amplitude of the vibration such that its magnitude subjectively fit the numbers that had been presented to them (magnitude production). These are both techniques that are often used in psychophysics. Figure 14 illustrates the results of the magnitude estimation and magnitude production procedures for a 25 Hz vibration.

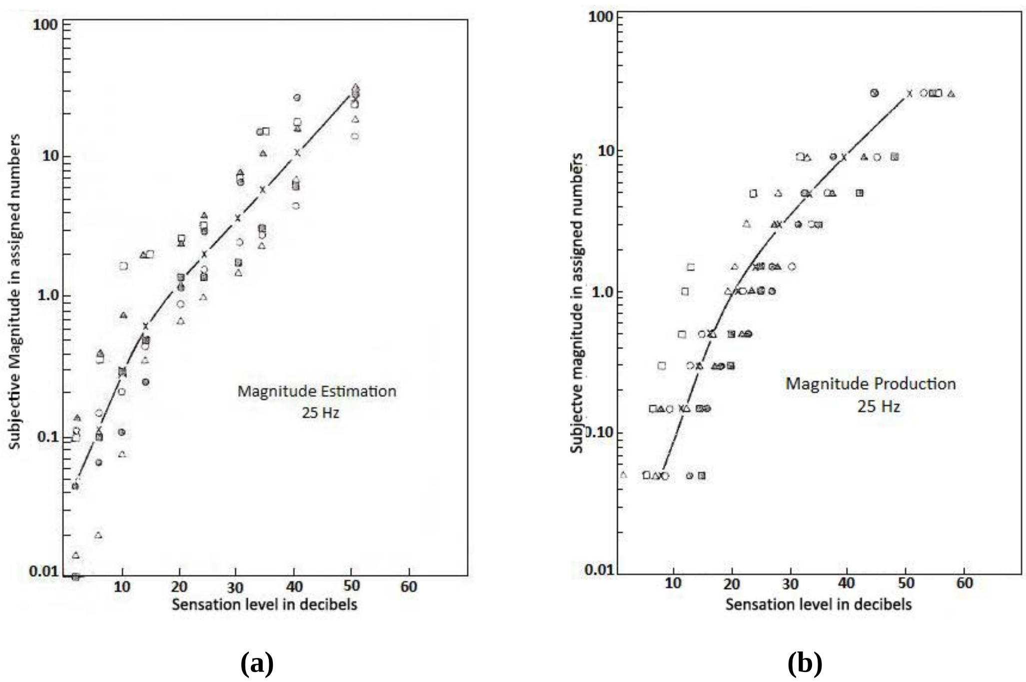
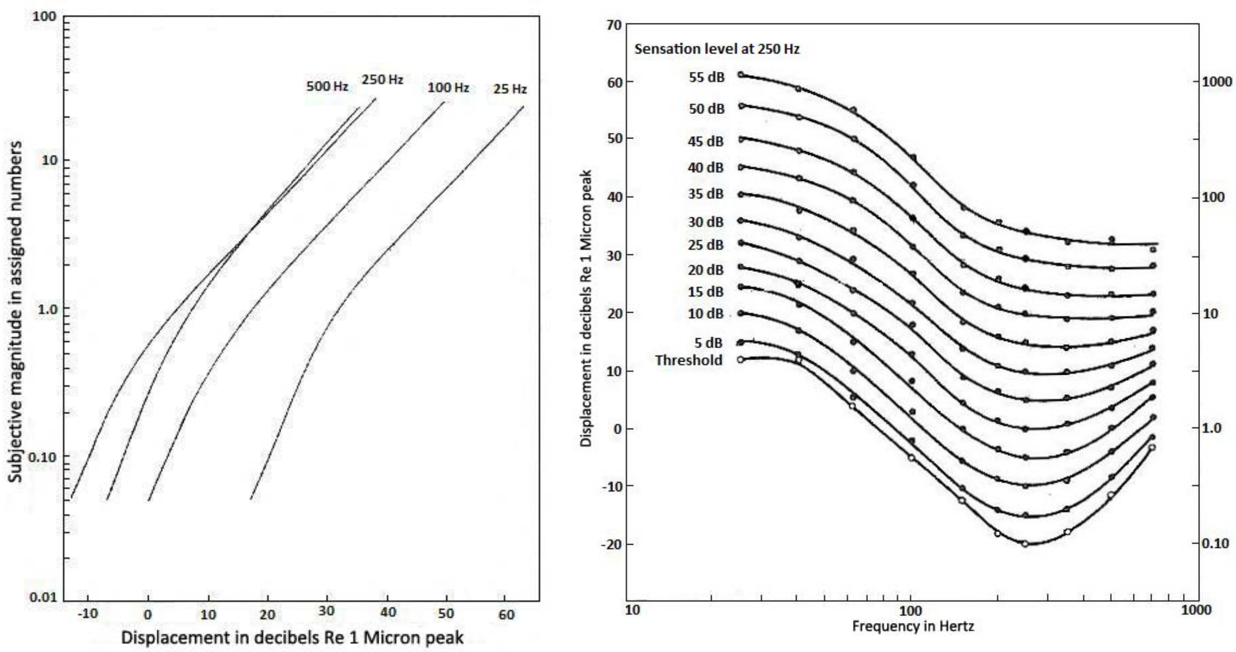


Figure 14: Results of magnitude estimation (a) and magnitude production (b) for a 25Hz vibration for different participants. Solid lines illustrate geometric means. Figures adapted from Verrillo et al. (1969, p. 368).

For each frequency tested, the geometric mean of the individual responses for the magnitude estimation and magnitude production functions were calculated. These functions were averaged, and curves of numerical magnitude balance were obtained. The curves in Figure 15(a) indicate that the perceived intensity of a vibratory stimulus at a given frequency grows as a power function of the stimulus amplitude. The exponents found for the power function were 0.89 for 25-300 Hz, 0.95 for 500 Hz, and 1.2 for 700 Hz vibrations. Stevens' findings (1968) also provide further evidence that the perceived intensity of a vibratory stimulus grows as a power function of stimulus amplitude. The slope of this function increases more rapidly on locations with lower sensitivity to vibration, such as torso. Taken together, this suggests that changes in the amplitude of a vibrotactile are perceived to be greater on the torso (Jones & Sarter, 2008).

All of the experiment results from both sections of the Verillo et al. (1969) were collected and re-plotted in terms of displacement as a function of frequency. The resulting set of curves is presented in Figure 15(b), and illustrates the contours of equal-sensation of magnitude. According to these curves, the intensity of a 250 Hz vibrotactile with specific amplitude can be identically perceived as a vibration at lower/higher frequency with higher amplitude.



(a) (b)

Figure 15: Subjective magnitudes as a function of absolute displacement (a), Contours of equal sensation magnitudes, the sensation level indications refer to a signal at 250Hz (b). Figures adapted from Verrillo et al. (1969, p. 370-371).

The results from the mentioned studies reveal the fact that there is a large interaction between frequency and amplitude of a vibrotactile stimulus. Therefore it is recommended only one of these parameters should change when using vibrotactile displays (Jones & Sarter, 2008).

3.4.3 Coding Information by using Different Durations of Vibrotactile Stimuli

Different durations of vibratory stimuli can also be used to encode information in vibrotactile displays. Summers et al. (1997) found that **performance for detecting increasing or decreasing frequency in a vibrotactile stimulus improves as stimulus duration is increased from 80 to 320 ms**. When vibrotactile stimuli are used to present a simple alert, the **preferred duration of tactile stimuli is between 50 and 200 ms**. **Prolonged vibrations are reported to be annoying for users** (Kaaresoja & Linjama, 2005). However, vibrations with different durations can be aggregated to provide rhythmic units (Brewster & Brown, 2004; Brown, Brewster & Purchase, 2005). Brown et al. (2005) provided three different rhythms by grouping pulses of different durations together. They used these rhythms to present three different types of messages. They reported that participants were able to correctly recognize the three message types with an average accuracy of 93%. Van Erp (2002) also suggests that when a single vibrator is used to encode information in a vibrotactile display, **the time between signals must be at least 10 ms**.

3.4.4 Coding Information using Different Locations for Vibrotactile Stimuli

A vibratory stimulus exerted to the trunk can be localized with relatively high accuracy and reliability (Cholewiak et al., 2004; Van Erp, 2005a). Therefore, arrays of vibrotactors can be used to support a number of spatial orientation applications, such as representing the location of an object relative to body, presenting directions in navigation systems, or as a counter measure for spatial disorientation (Van Erp, 2005a; Van Erp, Groen, Bos, & Van Veen, 2006). In general, observers are more capable of correctly localizing stimulus near the spine and navel on the torso. These points can serve as anatomical reference points (anchor points) for the trunk (Cholewiak et al., 2004). There is a bias between the actual tactor location and the responses of observers regarding the location of the stimuli. This bias is toward the midsagittal plane (toward the navel for the front of the torso and toward the spine for the back of the torso (Van Erp, 2005a).

The ability of participants to localize a vibrotactile stimulus in a 3×3 tactor array was investigated in an experiment by Linderman and Yanagida (2003). The vibrotactor array was affixed to the backrest of an office chair, such that vibrations were presented to the lower back region of the participant's torso. The spacing between the centers of each pair of neighbouring tractors was 6 cm. Lindeman and Yanagida found that participants were able to report the correct location of the tactors with an accuracy of 84% (Lindeman & Yanagida, 2003). In addition, they found that the spacing between tactors influenced the localization accuracy and must be adjusted carefully in design of vibrotactile displays. This is especially true when they are being used to convey spatial information. It is recommended that the inter-tactor spacing on the skin be greater than the two-point threshold for vibration (Jones & Sarter, 2008). As stated in Section 3.3**, the inter-tactor spacing on the trunk should be at least 3 cm for better localization performance** (Van Erp, 2005b)**.**

3.5 Vibrotactile Patterns

The current literature suggests that vibrotactile patterns can be classified based on the number of tactors used into two main groups: spatio-temporal patterns and *tactons*. Spatio-temporal patterns can be generated by sequentially activating a series of vibrotactors and require more than a single vibrotactor. Tactons, on the other hand, consist of a single vibrotactor and is manipulated by turning the tactor on and off. These two types of patterns are discussed in detail in the following subsections.

3.5.1 Apparent Movement and Spatio-Temporal Patterns

Spatio-temporal patterns and perceptions of apparent movement can be generated by sequentially activating a series of vibrotactors placed on the skin. Resulting patterns can be used to intuitively present information regarding orientation or direction of external events. Cholewiak and Collins (2000) investigated the influences of timing parameters and presentation modes on the generation of vibrotactile patterns. In this study, patterns were presented to the distal pad of the left index finger, the left forearm and the lower back region by means of seven vibrotactors for each area. Two modes of pattern presentation were used; saltatory and veridical. In veridical mode, all seven of the vibrotactors that were situated in a linear array were activated in sequence to provide a linear pattern. In salutatory mode, seven bursts of vibration were presented at only three tactor sites. Three bursts of vibration presented through the first; three bursts through the fourth; and one burst through the seventh vibrotactor. Figure 16 illustrates the concepts of these presentation modes. The distance between the adjacent tactors was 2.54 mm on the finger tip and 15.24 mm on the forearm and the lower back. The vibrotactors which were used for the fingertip were smaller in size than those were used for the forearm and the lower back. The vibrations were presented in the two modes with different *BDs* and *IBIs (Inter Burst Interval).* The values for the *BDs* and the *IBIs* were 4, 9, 17, 26, 35, 70, and 139 ms. The experiment was done in two main parts. The main goal of the first part of the experiment was to find out how efficiently a line could be generated. Participants were instructed to rate the levels of perceived length, smoothness, spatial distribution, and straightness of the presented patterns.

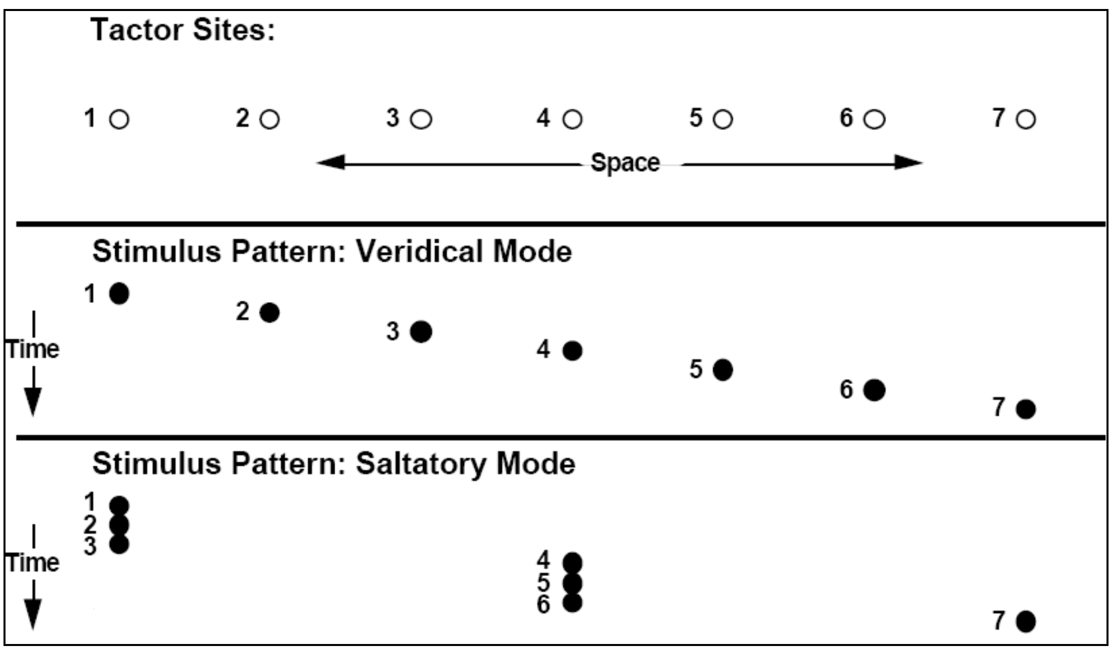


Figure 16: Concepts of veridical and saltatory presentation modes. Figure adapted from Cholewiak and Collins (2000, p.1223).

During the second part of the experiment, vibrations were presented only to the lower back. The aim of the second experiment was to find out to what extent are participants able to differential between the two presentation modes (veridical and salutatory), and which of these modes can generate a better sensation of a line.

The results of the first experiment showed that when vibrations were presented with longer BDs, participants perceived longer lines. Significant interaction between BD and IBI was also found. With longer IBIs for stimuli with a given BD, the generated lines were reported to have longer length. This means that as the velocity of activation sequence increases, the perceived length of the pattern decreases. The stimuli were also perceived to be smoother with shorter IBIs. Perceived smoothness of patterns was found to be mainly a function of IBI. Perceived spatial distribution was reported to have better quality when small BDs and IBIs were used. Finally, judgments of straightness improved with shorter BDs and shorter IBIs which indicates that increased velocity of an activation sequence will result in judgments of straighter patterns. This finding is in consent with the findings of Langford, Hall, and Monty (1973). A line produced by a moving point across the skin appears to wander at lower speeds and it is perceived to be straight at higher speeds (Langford, Hall, & Monty, 1973). The results of the second portion of the experiment revealed that the verdical mode was superior to the salutatory mode, but the differences were very small.

In addition to the apparent movement illusions explained above, the simultaneous activation of two vibrotactors located spatially close together causes the sensation of only a single point between the two tactors (apparent location). This point shifts continuously toward the vibration with higher intensity (Scherrick, Cholewiak, & Collins, 1990).

Kirman (1974) investigated the effects of *stimulus onset asynchrony* (*SOA*) and stimulus *burst duration (BD)* on vibrotactile apparent movement. Vibrotactile stimuli were presented to two different locations on the right index finger. The vibrations were varied in both duration and the inter-stimulus onset interval. They were presented in 6 durations (1, 10, 20, 50, 100, and 200 ms) and were combined with each of 10 SOAs (10, 20, 30, 50, 70, 90, 110, 130, 150, and 200 ms). Therefore a total of 60 pairs of stimuli were presented to the participants. Kirman (1974) found that the quality of perceived apparent movement varies as a function of SOA. Figure 17 shows this function for stimuli with durations of 200 ms. The best feeling of apparent movement for the two stimuli was achieved when the inter-stimulus onset interval was approximately equal to 130 ms. This means that the second stimulus started to stimulate 130 ms after the onset of the first stimulus. This also resulted in a 70 ms overlap between the two stimuli.

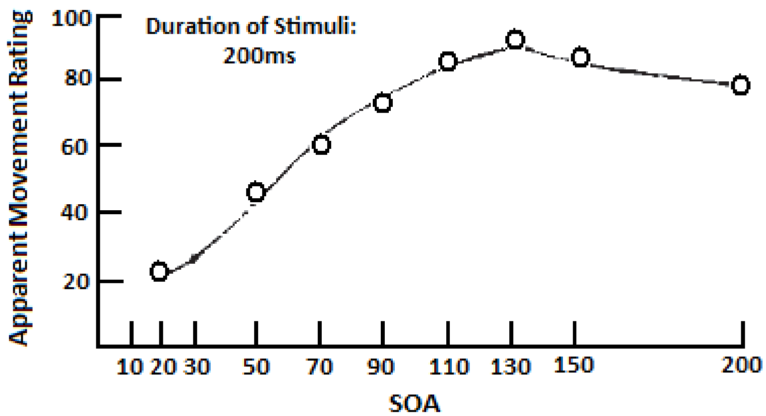


Figure 17: Apparent movement rating as a function of SOA. Figure adapted from Kirman (1974, p. 2).

Figure 18 shows the optimal *SOAs* for different stimuli durations applied in the experiment. As can be seen in the figure, participants were able to optimally perceive apparent movement when the *SOA* were 70, 50, 50, 70, 90, and 130 ms respectively for stimuli with durations of 1, 10, 20, 50, 100, and 200 ms respectively.

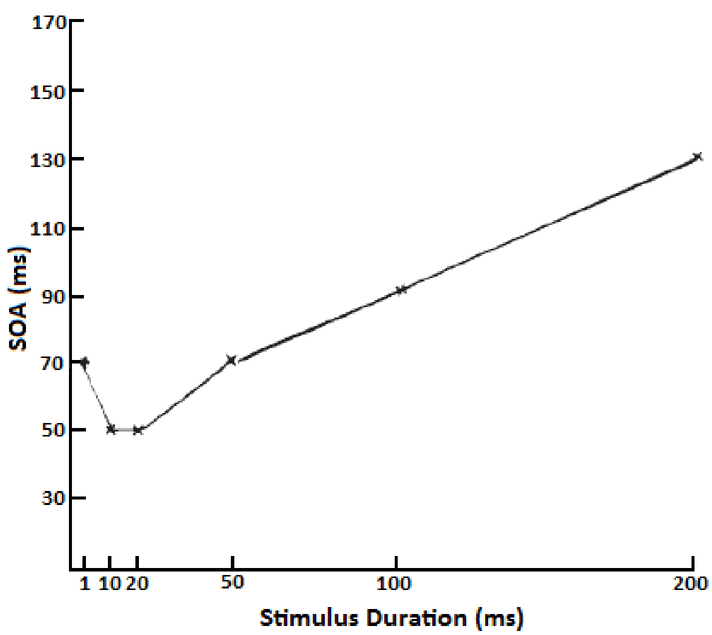


Figure 18: Optimal SOA as a function of stimulus duration. Figure adapted from Kirman (1974, p. 3).

Finally, Figure 19 shows the judgments of apparent movement for the optimal *SOAs* as a function of stimulus durations. According to this figure, as stimuli duration increases, judgments of apparent movement increase for optimal *SOAs.* Taken together, the results of this study suggest that **when spatio-temporal patterns are used in vibrotactile displays, the quality of perceived apparent movement is a function of inter-stimulus onset interval and *burst duration*.**

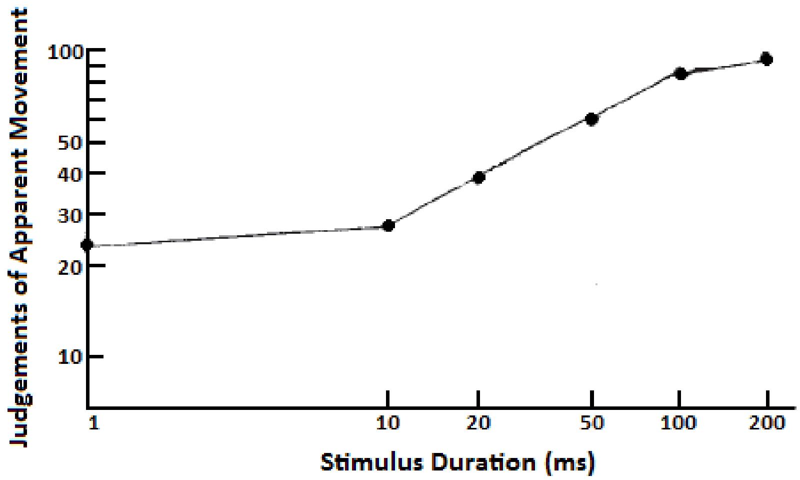


Figure 19: Judgments of apparent movement for the optimal SOAs as a function of stimulus duration (results of Kirman experiment). Figure adapted from Kirman (1974, p. 5).

While designing vibrotactile displays, it is important to remember that the number of patterns that can be generated is dependent on the number of arrays of vibrotactors embedded in the display. Jones, Lockyer, and Piateski (2006) presented navigational direction messages to participants through a set of vibrotactile patterns. The patterns were presented using a 4×4 tactor array mounted on the lower back of the participants. The participants navigated through a path designated by a grid of cones. Jones et al. (2006) found that participants were able to accurately follow the navigational commands to walk through the course using this aid. A visual depiction of one of the vibrotactile navigational commands is illustrated in Figure 20. Yanagida, Kakita, Lindeman, Kume, and Tetsutani (2004) investigated the participant's ability to recognize patterns which were used to present English letters and numbers. The patterns were presented through a 3×3 tactor array affixed to the backrest of an office chair. The sequential presentations of the patterns were such that they traced the trajectory in a manner that simulated hand writing on the back. Yanagida et al. (2004) found a ratio of 87% correct letter or number recognition. Although letter recognition was relatively successful in this experiment, it should be noted that in high workload conditions the accuracy may not stay the same. In the Ynagida et al. (2004) experiment, participants were not asked to perform any additional tasks beyond the recognition task. Thus, the workload for the participants was relatively low.

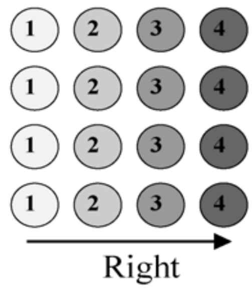
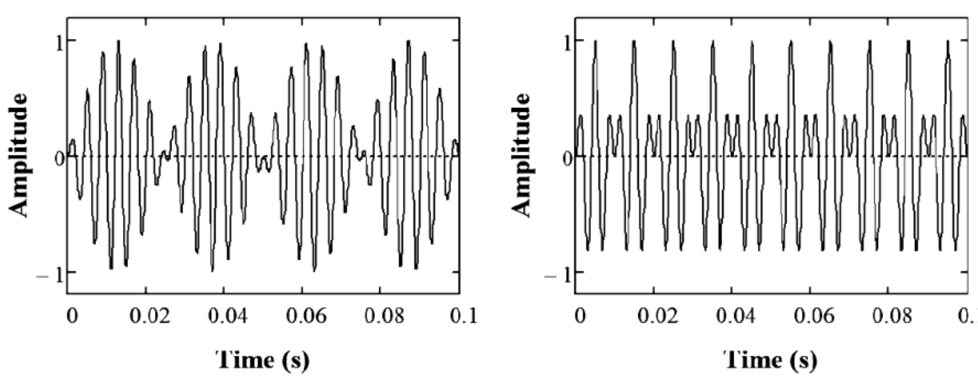


Figure 20: The pattern generated through a 4×4 array of vibrotactors for "turn right" command. The arrow represents spatial order of activation of tactors. Figure adapted from Jones et al. (2006, p. 1367).

It should be noted that the participant's familiarity with the displayed set of patterns may also affect the accuracy of the pattern recognition process. Therefore, practising may improve the discrimination performance for vibrotactile patterns (Gallace, Tan, & Spence, 2007; Yanagida et al., 2004).

3.5.2 Tactons

Vibrotactile patterns can also be generated by means of a single tactor. These patterns are called *Tactons*. *Tactons* are brief messages that can be used to represent complex concepts and information in vibrotactile displays. They are tactile replication of icons or *earcons* (Brewster & Brown, 2004; Brown et al., 2005). Brown et al. (2005) generated *tactons* by using different rhythms and waveforms. As mentioned previously, vibrations with different durations can be grouped together to create rhythmic units. Complex waveforms can be generated using sinusoidal amplitude modulation, as illustrated in Figure 21.



(a) **(b)**

Figure 21: 250Hz sine wave modulated by 20 Hz (a) and 50 Hz (b) sine waves. Figures adapted from Jones and Sarter (2008, p. 104).

The feeling of roughness can be transmitted by presenting participants with amplitude modulated signals through vibrotactors (Brown et al., 2005). Brown et al. (2005) found that participants are able to differentiate different amplitude modulated signals in terms of roughness. Sinewaves with no modulation are perceived as being smoother, and the feeling of roughness increases as modulation frequency decreases. Brown et al. (2005) preferred not to use vibrotactile parameters such as frequency or amplitude for creating *tactons*. The limited bandwidth of tactors and electrical devices discourage the use of different levels of frequency. Reducing the amplitude may make the pattern undetectable and increasing amplitude may cause pain and cause annoyance (Brown et al., 2005).

Brewster and Brown (2004) categorized *tactons* in three main groups; compound *tactons*, hierarchical *tactons* and transformational *tactons*. Brown et al. (2005) investigated the ability of a group of participants to identify different rhythms and different roughness levels when the characteristics are combined together to form transformational *tactons*. A single C2 tactor was used in the experiment. Three types of alerts (voice call, text message and multimedia message) were encoded using different rhythms. The priority of the alerts (low, medium or high) was encoded using different roughness levels. For example, the same rhythm was used to present a high priority text message and low priority text message, but they were presented with different roughness levels. Brown et al. (2005) found average discrimination rates of 93% and 80% for the different alert types (represented by different rhythms) and alert priority levels (represented by different roughness levels) respectively. The average result for overall *tacton* recognition was 71%. Considering these results, **we can conclude that in vibrotactile displays, *tactons* can effectively convey complex messages to the operators in a very concise manner.**

3.6 Other Reviews of Tactile Characteristics

Many other researchers have reviewed coding principles and characteristics of vibrotactile stimuli. One recent review by Self, Van Erp, Eriksson, and Elliott (2008) discussed nine tactile characteristics which designers may be able to manipulate in order to communicate messages. While many of these have been discussed above, a table (taken from Self et al. (2008) is included below to show some other possible methods of coding information into the tactile modality.

Table 3: Tactile Characteristics (Self et al., 2008, p. 4)

Characteristic	Properties
Size	• Limited number of distinctive levels • Large difference between sizes preferable • A clear boundary is needed • Simultaneously displayed sizes is feasible

Shape	• Fair number of distinctive levels • Similar tactile shapes should be avoided • A clear boundary is needed • Simultaneously displayed shapes is feasible
Orientation	• Many distinctive levels possible • Large distance between displays preferable • Simultaneously displayed positions is highly feasible
Position	• Many distinctive levels possible • Large distance between displays preferable • Simultaneously displayed positions is highly feasible
Moving patterns	• Any distinctive levels possible • The moving patterns should be quickly recognizable after their start • Simultaneously displayed moving patterns is moderately feasible
Frequency	• Limited number of distinctive levels • Low feasibility for simultaneously displayed frequencies
Amplitude	• Limited number of distinctive levels • Low feasibility for simultaneously displayed amplitudes
Rhythm	• Many distinctive levels possible • The rhythms should be quickly recognizable after their start • Low feasibility for simultaneously displayed rhythms
Waveform	• Includes square, triangular, saw tooth, and sine waves • Requires sophisticated hardware

3.7 Masking Effects

Masking occurs when two stimuli are presented close to each other in space or time and decrease the detectability of each other. It is the difference between the perception of a stimulus when it is presented solely, and the perception of the same stimulus when it is presented close to another stimulus, either in time or space. In the design of vibrotactile displays, masking effects can play a large role in how operators perceive the messages. It is possible that an operator might miss an important piece of information due to masking by a nearby tactor, especially if multiple streams of data are presented through the vibrotactile display.

In general, we use the term "target" for the stimulus which is to be identified, and the term "masker" for the stimulus which is to be ignored. The masker stimulus may change several discriminating parameters of the target stimulus (e.g. sensation threshold, difference threshold and perceived location of the stimulus (Cheung, Van Erp, and Cholewiak, 2008; Craig & Evans, 1987).

3.7.1 Temporal Masking

Temporal masking occurs when the vibrations are presented to the same location, and the target stimulus is presented either within the time interval of the masking stimulus, or near the onset or just after the offset of the masking stimulus. Temporal masking decreases when the temporal separation between the onsets of stimuli increases (Van Erp, 2002; Cheung, Van Erp, and Cholewiak, 2008). *Forward masking* occurs when the target stimulus is corrupted with a preceding masking stimulus. *Backward masking* occurs when the target stimulus is corrupted with a subsequently presented masking stimulus. Participants are better able to recognize tactile patterns when they are presented in isolation than when they are presented with a forward or backward masker. Higher masking levels occur at shorter *SOAs* (Craig & Evans, 1987). Craig and Evans (1987) presented a masker pattern followed by a target pattern to participants who were instructed to identify the second pattern while ignoring the first pattern. They found that with shorter *SOAs* there was more backward masking than forward masking. As *SOAs* increased, forward masking decreased more gradually than backward masking. Craig and Evans (1987) also reported that with long *SOAs*, the opposite was true and there was more forward than backward masking. Forward masking remained visible for *SOAs* up to 1200 ms.

In another study, Gescheider, Bolanowski, and Verrillo (1989) investigated the amount of simultaneous, forward, and backward masking. In this experiment a 700 ms vibratory stimulus was used as the masker, and a 50ms vibration was used as the target. The *SOA* was varied over a range of 2000 ms. The target stimulus was presented within the time interval of the masking stimulus (simultaneous masking), presented with partial overlap with the masking stimulus, or without any overlap with masking stimulus (forward and backward masking). The effect of temporal masking was strongest when the target stimulus was presented near the onset or just after the offset of the masking stimulus. The amount of masking declined as the time interval between masking and target stimuli increased. The rate of decline of the masking effect appeared to be same for forward and backward masking. Despite the findings of Craig and Evans (1987), they did not report the persistence of forward masking for long *SOAs*. This difference between the results is probably due to the different methodologies used. As mentioned previously, Craig and Evans (1987) used patterns of vibration in the form of vertical or horizontal lines as stimuli, whereas Gescheider et al. (1989) used single vibrations as stimuli. Unfortunately, **we cannot make any strong conclusions regarding temporal masking** based on the current literature.

3.7.2 Spatial Masking

Spatial masking occurs when two stimuli are presented to two distinct locations at different or overlapping times (Van Erp, 2002; Cheung, Van Erp, and Cholewiak, 2008). We can reduce the amount of spatial masking by increasing the distance between stimulated sites (Cheung, Van Erp, and Cholewiak, 2008; Cholewiak, Collins, & Brill, 2001). When stimuli are presented at different times, spatial masking occurs only when the target and the masker stimuli are both high frequency vibrations. Therefore the effect of spatial masking is greater on receptors within the Pacinian system. Non-Pacinian systems do not demonstrate this characteristic, unless the stimuli are presented at the same time (Verrillo & Gescheider, 1983). Craig (1974) measured the difference threshold in the presence and absence of a masking stimulus. When the difference threshold was measured in the presence of the masking stimulus, a masking vibration was presented simultaneously with the test stimulus. The test stimulus was a 160 Hz vibration presented to the right index finger. The masking stimulus was a vibration with the same frequency delivered to the right little finger. The results of this experiment demonstrated that the difference threshold of the target stimulus considerably increases as the intensity of the masker stimulus increases. Only when the intensity level of the target stimulus was more than 15 dB above threshold, the difference threshold in the presence of the masking stimulus was similar to the difference threshold in the absence of masking stimulus. In order to reduce the negative effects of spatial masking, **it is recommended that vibrotactors which have a static surround in their structure should be used** (e.g. C2 tactors). A rigid surround can prevent the spread of vibrations and surface waves to adjacent locations reducing the effect of spatial masking (Van Erp, 2002; Cholewiak et al., 2001).

3.8 Tactile Perception Summary

Before we move onto addressing crossmodal attention and examining how tactile displays fit in the operator’s perception of a multisensory environment, we summarize the findings discussed. The art of designing vibrotactile displays is still in its infancy. Currently, one important focus in the design of such displays is their capability in navigation tasks in 3D space. The three-dimensional nature of the torso can facilitate the understanding of three-dimensional spatial information. Most researchers who have investigated the localization ability and spatial acuity of the skin for vibratory stimuli have used a single array of vibrotactors. There are relatively few studies which have examined these abilities while using multiple rows of tactors. Other uses of tactile displays such as alerts and other methods for coding other types of non-spatial information are also actively being explored. Human factors issues have major influences on design and application of any vibrotactile display. Therefore, we should consider the perceptual factors in pattern generation and coding procedures used to design future vibrotactile displays. Relevant guidelines for different ways of information presentation on these displays are provided in this section.

We can code information by presenting vibrations with different frequencies, amplitudes, durations, and locations on the body. Optimal sensitivity of human skin to vibration is between 150 to 300 Hz. The detection threshold as a function of frequency is a U-shaped curve which has its minimum in the region of 250 Hz. High levels of interaction between frequency and amplitude of a vibrotactile stimulus suggest that only one of these parameters should be changed for coding information. Also, there is a high level of uncertainty about the perception of change in frequency by human skin.

Changes in the amplitude of vibration can be perceived with relatively good accuracy which makes it a very useful parameter to encode information in vibrotactile displays. Vibrations with different amplitudes can be used to create different levels of intensity.

Different durations of vibratory stimuli can also be used to encode information. When a vibrotactile stimulus is being used to present a message, the duration of vibration should be between 50 to 200 ms. Prolonged vibrations are annoying for users. Also, vibrations with different durations can be grouped together to provide rhythmic units which can be used to generate *tactons*.

A vibratory stimulus exerted to the trunk can be localized with relatively high accuracy and reliability. This fact makes the location of a vibration an important parameter for coding information in vibrotactile displays. In general, observers are more capable of correctly localizing stimulus near the spine and the navel on the torso. These points can serve as anatomical reference points (anchor points) for the trunk. We should consider taking advantage of these anatomical points of reference for coding information in a vibrotactile torso display. For better localization performance, the inter-tactor spacing on the skin should be greater than the two-point threshold for vibration. For the trunk, the inter-tactor spacing should be about 3 cm.

Based on the number of tactors employed to represent messages in a tactile display, vibrotactile patterns can be divided into two main groups: *tactons* and spatio-temporal patterns. *Tactons* can effectively convey abstract messages to the operators by means of a single tactor. Spatio-temporal patterns can be generated by sequentially activating a series of vibrotactors and can be used to intuitively present information regarding orientation, direction, or more abstract concepts. Obviously, the number of distinctive patterns that can be generated through a vibrotactile display is dependent on the number of arrays of tactors in the display. Therefore, spatio-temporal patterns provide a larger set of possible discriminable patterns than tactons.

It should also be noted that when information is being presented through a vibrotactile display, all of the tactors must have proper contact with the skin, such that the vibrating contractor (the part of the tactor that makes contact with the skin) maintain contact with the skin. Otherwise, part of the message may be missed or a tactile pattern may be incorrectly perceived as a different but similar pattern.

4 Auditory Display Design and Presentation of Urgency Information

As mentioned previously, both the visual and auditory modalities have been the focus of a large body of research. Compared to the tactile modality, the research on the auditory modality has progressed to a stage where researchers are now able to focus on adapting auditory perception research to the design of effective alerts and displays in real-world problems. In a review of the role of psychoacoustics research on the design of auditory displays, Walker and Kramer (2004) state that the task of interacting with auditory display can be described with three subtasks: hearing, grouping, and meaning making. Hearing refers to the basic perception of auditory stimuli. Research in this area is focused on how well individuals are able to perceive auditory information, and this subtask covers many of the topics covered in the tactile perception section of this report, such as detection thresholds, discrimination sensitivities, and masking (Walker & Kramer, 2004). This knowledge serves as the foundation for the other higher level tasks, and within the auditory domain this research has already been well established (see Neuhoff, 2004 for a review of ecological psychoacoustics), in contrast to the ongoing debate that still exists in the tactile domain.

The tasks of grouping and meaning making are issues that must be dealt with in the design of complex displays. Grouping refers to how individuals are able to parse incoming stimuli into channels or streams of data (Walker & Kramer, 2004). Topics such as the cocktail party effect (Arons, 1992) and auditory scene analysis (Bregman, 1990) are directly related to this task. Meaning making, on the other hand, refers to the cognitive processes that occur when an individual attempts to relate the perceived stimuli to meaning. This task represents a key difference between auditory displayer designers (as well as designers of any sensory modality) and psychophysics researchers. In the design of displays, the focus is not on how an individual perceives a physical stimulus; instead the focus is on how well the individual is able to interpret the physical stimulus with respect to the information that the display designer is attempting to communicate.

Therefore, the focus of this section will be on the use of auditory stimuli in displays. In particular, we will describe different methods for coding information into auditory messages that can be discriminated by users. We also provide and discuss in detail how urgency information has been encoded into auditory information. Where possible, comparisons to the visual and tactile modalities will be discussed.

This section is organized as follows:

- Section 4.1. Describes different auditory coding methods, and provides insight into the benefits and drawbacks of the various methods.

- Section 4.2. Provides a discussion of current ways that urgency information is coded in the auditory, visual, and tactile modalities.

- Section 4.3. Provides concluding remarks.

4.1 Coding Methods Within the Auditory Modality

As mentioned previously in Section 2, there are a number of methods that designers have used to code information into the auditory modality. Sanderson and Watson (2005) included *earcons*, *auditory icons*, *audification*s, and *sonifications* as examples of methods in their discussion of adapting EID for usage with designing auditory displays. These methods, along with others will be described in further detail within this section. For a coding method to be successful, the listener must be able to extract the required information from the physical auditory stimulus. This is normally accomplished through the perception of different attributes of physical stimulus, such as frequency (which is perceived as pitch), volume (loudness), tempo and rhythm (which describes the speed, rate, or frequency of a auditory of event), and timbre ("a catch-all term...used to mean all those sound attributes that are not loudness, pitch or tempo." (Walker & Kramer, 2004, p. 159)). However, Walker and Kramer (2004) also state that the context of the signal (e.g. the environment, tasks to be accomplished, etc.) also plays a large role in how the stimuli is understood.

4.1.1 Dimensions for Categorizing Auditory Coding Methods

The designer of the auditory display must decide on the complexity of message which they are attempting to communicate as well as how this message is semantically mapped into the sound characteristics. Walker and Kramer (2006) established a taxonomy of auditory coding methods based on a *symbolic-analogic continuum*. They describe symbolic displays as ones that "establish a mapping between a sound and an intended meaning, with no intrinsic relationship existing." (p. 1022) In contrast, analogic displays "contain an immediate and intrinsic relationship between the display dimension and the information that is being conveyed." (p. 1022) Figure 22 provides examples of different methods of auditory coding along this continuum.

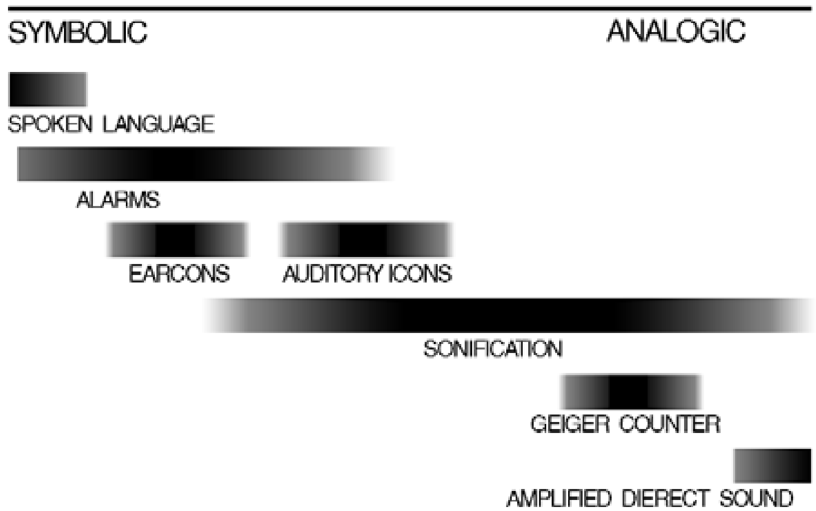


Figure 22: The symbolic-analogic continuum with examples of different types of auditory coding methods. Taken from Walker and Kramer (2006, p. 1022).

Research by Stephan, Smith, Martin, Parker, and McAnally (2006) has shown evidence that auditory displays that employ more analogic forms of coding are easier to learn and remember. The authors were interested in studying how well participants were able to remember the pairings of different auditory icons and events between auditory icons and events. Auditory icons are *auditory signals* that have a strong analogic link to an object or process. For example, the sound of a door closing is often used to signify the process of someone leaving a chatroom in online chatting applications. However, Stephan et al. (2006) noted that not all auditory icons employ the same degree of association between the signal and its referent. Some auditory icons make direct references to its referent (e.g. the association between the sound of a dog and concept of a dog), while other auditory icons make only indirect references to its referent (e.g. the association between the sound of a seagull and the beach). Therefore, the authors tested three different strengths of association: direct, indirect, and unrelated.

Participants were asked to learn pairings between auditory icons and then recall them after 4 weeks. Stephan et al. (2006) found that the pairings which were unrelated, and thus was the most “symbolic”, led to the greatest number of recall errors both right after learning the pairings, and after the four week interval. The indirect and direct pairings, however, did not differ in terms of their learnability (performance immediately after the training session). However, after an interval of four weeks participants were significantly better at recalling direct pairings than indirect pairings. The authors also found support that stronger associations between the signal and referent lead to faster processing of the auditory icon. This supports evidence found by Belz, Robinson, and Casali (1999) that found auditory icons out performed traditional auditory warnings. **Taken together, these findings suggest that as the degree of analogy increases in an auditory display, participants are able to better remember and more quickly process the auditory stimuli.**

A second continuum, while not explicitly described by Walker and Kramer (2006), also exists that varies based on the complexity of the message being coded and communicated. At one end of the continuum are simple auditory alarms and notifications, which indicate the presence or absence of an event often with the goal of capturing the user's attention (Walker & Kramer, 2004). Alarms tend to only present binary information, hence they reveal very little about the event that caused the notification. Additional information can be coded into the auditory stimuli by increasing the complexity of the signal. At the other end of the continuum lie sonifications which are complex auditory stimuli that "transform data relations into auditory relations." (Walker & Kramer, 2006)

A sonification, designed by Watson and Anderson (2000), for assisting with autolanding commercial aircraft was discussed in Section 2. This sonification mapped task relevant information such as air speed and direction of thrust onto auditory characteristics, tempo and pitch respectively. The tempo of the auditory carrier signal would increase as the air speed of the aircraft increased, thereby communicating changes in the data through changes in the auditory stimuli. However, it becomes much more difficult to find effective methods for mapping data onto auditory characteristics as the message becomes increasing complex. Pollack and Ficks (1954 as cited by Walker & Kramer, 2004) found that auditory displays which varied multiple auditory displays did not perform as well as auditory displays that made use of a single auditory characteristic. Walker and Kramer (2004) recommend that one method for improving the performance would be to map a single set of a data onto a set of auditory characteristics.

Furthermore, research conducted by Walker and Kramer (1996) has found that different types of data may have auditory characteristic mappings that are more intuitive for a user who is attempting to make sense of the display. In the experiment to investigate how the conceptual understanding of a data type is affected by the type of auditory characteristic used, participants were trained to associate temperature, pressure, size, and rate information with loudness, pitch, tempo, and onset sharpness. Similarly to the Stephan et al. (2006) experiment described above, the mappings (data type and auditory characteristic) were varied across participants. However, the sonification represented a single "sound" that was composed of multiple dimensions, whereas the auditory icons used in Stephan et al. (2006) only made use of a single signal-referent pair for each sound. After being trained on these associations, participants were asked to monitor signals and respond accordingly when one of the data parameters deviated from normal. For example, when the temperature variable dropped, participants were required to press a heater button.

Walker and Kramer (2005) had predicted that some of the mappings would be intuitive, and therefore lead to the highest accuracy and response time. The intuitive pairings were temperature with pitch, pressure with onset, size with loudness, and rate with tempo. Surprisingly, the authors found that their hypothesized ideal mappings were completely incorrect, and quite often the predicted "bad" or "random" pairing would actually produce the fastest and most accurate responses. They concluded that **even with training, some participants still showed preferences for mappings between certain types of data and certain auditory characteristics.** This is a point that will be further discussed in the following sections on urgency presentation. A second finding that Walker and Kramer (2005) state is **that the polarity of a mapping (the direction that the auditory characteristic changes whenever the input data changes) is also an important design element to consider**. They used the reverse polarity of increasing mass mapped onto decreasing pitch as an example. Further research on this topic has shown that

different groups of listeners may also have different underlying assumptions about the intuitiveness of a mapping (Walker & Lane, 2001 as cited by Walker and Kramer, 2004). Overall, this line of research has highlighted the importance of careful semantic mapping choices, especially with complex auditory displays.

4.1.2 Auditory Coding Methods

The symbolic-analogic and complexity continuums allow us to classify different methods for coding auditory information. Table 4 lists the most common auditory coding methods described in the literature, ranked from the most symbolic to the most analogic. For comparison, a list of similar coding methods in the visual and tactile modalities is provided. Some of these coding methods, especially in the tactile modality, are not yet formally defined and are purely speculative at this point. The purpose of developing this table is to depict possible equivalencies of coding methods across modalities. Such equivalencies can serve as a possible guideline in the future when encoding messages in the vibrotactile display.

Table 4: Comparison of Coding Methods for Audition, Vision, and Touch

Audition	Vision	Touch
Earcons: "a discrete sound that is a member of a set of sounds that are related to each other through a syntactic structure" (Sanderson & Watson, 2005). Earcons tend to make use of generic tones that rely heavily on the symbolic link between the tone and a concept. *Example: "A three-note pattern representing a file, in which a decrease in loudness and pitch represents "file deletion" – the diminishing loudness and pitch of the sound is a metaphor for the destruction of the file." (Walker & Kramer, 2004, p. 152)*	**Analogous Icons:** an icon that visually captures a constraint in the environment. (Burns & Hajdukiewicz, 2004). *Example: A map captures spatial relationships and visually depicts them.*	**Tacton**: a brief tactile message that can be used to represent complex concepts and information in a vibrotactile display. *Tactons* can be generated by exerting different rhythms and waveforms to a single tactor (Brewster & Brown, 2004; Brown, Brewster, & Purchase, 2006a). *Example: Different Types of alerts (e.g. voice call, text message) can be encoded using different rhythms of a single tactor. (Brewster & Brown, 2004)*
Auditory Icons: sounds that represent a thing that draws heavily from its real-world equivalent (Sanderson and Watson, 2005) *Example: The sound of a door closing to signify a person*	**Icons**: graphic symbols that represent a concept or process due to the similarities between the graphical element and its real-world equivalent (Burns & Hajdukiewicz, 2004).	**Ecological valid tactile patterns:** tactile stimuli that produces an easily recognizable real-world sensation. Not a formal term, and has not be explored in detail within the literature.

leaving a chatroom.	*Example: Small pictograms used in Microsoft Windows.*	*Examples: Vibrations generated by a pair of vibrotactors located on the left and right side of the body to monitor imbalance in a vehicle.*
Sonification: the mapping of a source or multiple sources in the world into auditory dimensions of an auditory signal (Sanderson & Watson, 2005). *Example: Geiger counter.*	**Data Visualization:** "an image constructed to convey information about data" (Keller & Keller, 1993) *Example: Polar star diagrams.*	**Spatio-temporal tactile patterns:** a pattern created by the sequential activation of a series of vibrotactors to intuitively present information using multiple dimensions. *Example: By sequentially activating a horizontal array of vibrotactors from right to left, a "left turn" concept can be generated (Jones, Lockyer, & Piateski, 2006).*
Audification: a translation of some physical stimuli into an auditory representation (Sanderson & Watson, 2005). *Example: Guitar amplifier.*	**Signal visualization:** a translation of some physical stimuli into a visual representation. *Example: Voltage or amplitude on an oscilloscope display.*	**Tactification:** a translation of some physical stimuli into a vibro-tactile representation. This is not a formal term, and has not been studied in detail in the literature. *Example: Seismic data presented through a tactor.*

Recently, there has been some investigation into the design of crossmodal coding methodologies. Hoggan and Brewster (2007) examined the design of audio and tactile crossmodal icons for use with mobile devices. By taking advantage of the fact that some coding methodologies, such as earcons and tactons, are highly related (similar position along the symbolic-analogic and complexity continuums) the authors designed messages that were similar in both sensory modalities. The authors termed these new messages crossmodal icons because they existed in similar forms across different modalities. This was accomplished through the use of sensory characteristics that were "amodal" and were similar in each modality. Hoggan and Brewster (2007) stated that intensity, rate, rhythmic structure, and spatial location were all examples of common characteristics shared between auditory and tactile stimuli. They tested their crossmodal messages by training participants in one modality and then testing them in another modality. For example, some participants were trained on the auditory version of the crossmodal icon (an earcon), and then they were tested using the tactile version (a tacton). A control group was trained and tested using messages from the same modality. Hoggan and Brewster (2007) found evidence that participants who were trained in one modality could translate this knowledge into understanding a similar icon in another modality. Participants achieved 85% accuracy when trained with earcons and tested on tactons, and 76.5% accuracy when trained with tactons and tested with earcons. The authors also found evidence that certain amodal characteristics were more effective in crossmodal icons. They found that roughness (achieved by modulating the amplitude of the signal) was not as effective as rhythm and spatial location. **Taken together, these results suggest that coding methods in different modalities with similar symbolic-analogic and complexity requirements can be designed using similar techniques.**

4.2 Urgency

Urgency and high priority levels are an important aspect to address in the design of interfaces. In cases of emergencies, the operator needs to be notified of the situation in the most effective manner. This requires that the operator understand that the incoming signal is relevant and important to their tasks and goals. Urgency is one example of a data type that auditory interface designers may find pertinent to encode into their displays because it is applicable to a large range of events. Mapping the perceived urgency of the alarm to the urgency of the situation is called urgency mapping. Urgency mapping is very essential in design of alarms and warnings. “It allows alarms to be matched meaningfully to the situations that they indicate and ensures that warnings contain information about their level of priority.” (Hellier & Edworthy, 1999)

In one example of the importance of presenting urgency, Ho, Nikolic, and Sarter (2001) conducted a study that examined the effectiveness of presenting operators with urgency information to support interruption management. Participants were required to handle interruption tasks that were presented through different modalities (vision, auditory and tactile). One group was given information about the interruption task in terms of urgency, time required to complete the task, and modality of the task (called the abridge group*)* whereas the other group was only given information about the presence of a pending task (called the basic group). Overall, the results in this study indicated that presenting participants with information about the urgency of the task, helped operators manage interruptions and as a result improved their task performance. Participants in the basic group performed significantly worse than those in the abridge group with high priority interruption tasks (Ho et al., 2001). This study demonstrates how essential urgency implementations are. In this section we explore some current implementations of urgency information.

4.2.1 Auditory Urgency

Traditionally, alarms have been used as one method for communicating the urgency of an event (Walker & Kramer, 2006). Auditory alarms, in particular, have been used in many applications and the users of these alarms range from specially trained pilots and nuclear power plant operators to individuals who have little or no training at all (e.g. individuals required to evacuate after hearing a fire-alarm). Thus, it is important that alarms, and other warning messages, are able to communicate urgency information in an intuitive manner.

Perceived urgency of auditory alarm is the impression of urgency that a listener gets when listening to a particular sound (Hellier & Edworthy, 1999). Perceived urgency can be modified by varying the acoustic properties of an auditory signal. Haas and Edworthy (1996) found that auditory signals which are rapid, and have shorter inter-pulse intervals, are perceived to have higher urgency. Haas and Edworthy (1996) also found that signals with higher intensity lead to higher perceived urgency, while higher frequency lead to faster response times. An applied example of perceived urgency in auditory alarms can be found in work done by Arrabito, Mondor, and Kent (2004). They conducted an investigation of the perceived level of urgency of auditory alarms in the CF CH-146 Griffon helicopter for trained CH-146 Griffon pilots and non-pilots. Arrabito et al. (2004) found that even with trained CH-146 Griffon pilots, the perceived level of urgency, as rated by the participants, did not match the urgency of the situation that the alarm represented. The mismatch between the perceived level of urgency and the urgency of the situation was even more pronounced for participants who were not trained pilots. Arrabito et al. (2004) found that properties of the auditory stimuli (such as frequency composition, repetition rate, amplitude, and harmonic relation of the frequency components) are intuitively interpreted by participants as being indicative of urgency. Specifically, auditory alarms that made use of multiple frequency components and a regularly modulated intensity invoked the highest perceptions of urgency. These findings are similar to those found by Haas and Edworthy (1996). However, these attributes were not always included in alarms that signified high priority events. In fact, alarms which were composed of relatively more continuous auditory signals (similar levels of signal ampltidue throughout) were rated as being less urgent. Because of these findings, Arrabito et al. (2004) concluded that the auditory alarms used within the Griffon helicopter were not adequately designed for their intended purposes.

The concept of perceived urgency was also examined by Hellier and Edworthy (1999). They recommended that non-verbal auditory alarms should be constructed such that they can present different levels of urgency. If this recommendation is followed, then auditory alarms with three or more levels of urgency (e.g. low, medium and high urgency) could be constructed. Consequently, alarms with different levels of urgency can be used to indicate different situations and conditions. For example, less critical conditions can be presented by less urgent alarms. The perceived urgency of an auditory alarm can be manipulated by varying the acoustic and temporal parameters of the alarm. For example, increasing an acoustic parameter such as pitch or a temporal parameter such as speed (decreasing the time interval between two pulses of sound) increases the perceived urgency of an alarm.

Hellier and Edworthy (1999) made use of *Steven's power law* to describe the relationship between changes in an objective parameter of an auditory alarm (e.g. pitch or speed) and the subjective perception of the urgency;

$$S = KO^m \tag{2}$$

Where:
S is the value of the subjective parameter.
O is the value of the objective parameter.
K is a constant.
m is the slope of the power function.

Steven's power law can be used for urgency mapping. The slope of this power function (m) indicates the magnitude of the relationship between the objective alarm parameter and the perceived urgency. For example a small change in an alarm parameter with a large exponent (m) provokes a large change in perceived urgency. To obtain this variable (m), subjective perceptions of urgency along various alarm parameters should be investigated by performing experiments.

Exponents of Steven’s power function for different alarm parameters such as pitch, speed, repetition rate, inharmonicity and length are determined and provided in Table 5.

Table 5: Steven’s power function exponents for five alarm parameters. Adapted from Hellier and Edworthy (1999, p. 170).

Alarm Parameter	**Definition**	**Exponent**
Pitch	Frequency of the auditory alarm	0.38
Speed	Pulse rate of the auditory alarm in a unit of sound	1.35
Repetition	Number of repetitions of a unit of sound	0.50
Inharmonicity	Number of inharmonic partials between the fundamental frequency and the first harmonic	0.12
Length	The total duration of the alarm in ms	0.49

Hellier and Edworthy (1999) found that speed is the most influential parameter of perceived level of urgency of an auditory alarm. Much larger changes in inharmonicity are needed to provide a unit change in perceived urgency. This was reflected in the exponent values found for speed and inharmonicity.

More recently, auditory alarms have been from sequences of notes with different pitches through a specific rhythm. Different levels of urgency can be indicated by playing the notes at different speeds. Sanderson, Wee, Seah, and Lacherez (2006) stated that higher levels of urgency can be indicated by playing the notes more rapidly, increasing the overall tempo of the auditory signal. In another study conducted by McNeer et al. (2007) auditory alarms with different structures were presented to participants who were required to judge the perceived urgency level of the various alarm sounds. The different sounds were categorized into three groups: harmonic interval sounds, melodic interval sounds and duty cycle sounds. The harmonic interval sounds consisted of two tones which were played at the same duration and had the same onset; the harmonic interval was varied within this group. The melodic interval sounds also used two tones, but the same tones were always used, instead the onset time of the second tone was varied. Finally, the duty cycle sound consisted of a single tone which was repeated five times with different durations and onsets. Multiple sounds for each category were created and tested by participants to estimate the range of perceived urgency that could be invoked. Harmonic interval sounds were able to represent the largest range of perceived urgency, (35-80%). The range of urgency was smallest for the melodic interval sounds (52-72%) and finally the urgency levels for the duty cycle sounds ranged from 38% to 70%.

An international standard referred to as IEC 60601-1-8 proposes a set of melodic alarms that can be used in medical electrical instruments to represent a range of medium and high emergency level alarms. Table 6 shows the structure of these alarms in the IEC 60601-1-8 standard which presents medium priority alarms by playing a pattern of 3 tone pulses once, and presents high priority alarms by playing 5 tone pulses played twice. Sanderson et al. (2006) reviewed evaluations made by several research groups regarding the effectiveness of these standards in presenting urgency levels and concluded that the proposed melodic alarms in this standard are "difficult to learn and easily confused." The IEC 60601-1-8 standard uses the same rhythms and number of tones for various types of alarms, making it difficult for users to understand and interpret. Thus, **if urgency is conveyed through the auditory modality, urgency codings should be as intuitive as possible so users can interpret them with as little effort as possible**.

Table 6: Sanderson et al. (2006, p. 25) presented a description of the melodic alarms proposed in IEC 60601-1-8 standard. The total duration for the medium priority alarms is approximately 920 ms and for each repetition of the high priority alarm is 1250 ms.

Alarm	Melody* and mnemonic lyric		Rationale mnemonic (other information in support of mapping)
	Medium priority	High priority	
General	C4-C4-C4	C4-C4-C4—C4-C4 (repeated)	Fixed pitch, traditional (usual) ISO 9703 sound
Oxygen	C5-B4-A4 "OX-Y-GEN"	C5-B4-A4—G4-F4 (repeated) "OX-Y-GEN A-LARM"	Slowly falling pitches; top of a major scale; falling pitch of an oximeter
Ventilation	C4-A4-F4 "VEN-TI-LATE" "RISE-AND-FALL"	C4-A4-F4—A4-F4 (repeated) "VEN-TI-LATE A-LARM" "RISE-AND-FALL AND-FALL"	Old "NBC chime;" inverted major chord; rise and fall of the lungs
Cardiovascular	C4-E4-G4 "CAR-DI-AC"	C4-E4-G4—G4-C5 (repeated) "CAR-DI-AC A-LARM"	Trumpet call; call to arms; major chord
Temperature (or delivery of energy)	C4-D4-E4 "TEM-P'RA-TURE"	C4-D4-E4—F4-G4 (repeated) "TEM-P'RA-TURE A-LARM"	Slowly rising pitches; bottom of a major scale; related to slow increase in energy or (usually) temperature
Infusion (drug delivery)	C5-D4-G4 "IN-FU-SION"	C5-D4-G4—C5-D4 (repeated) "IN-FU-SION A-LARM"	Jazz chord (inverted 9th); drops of an infusion falling and "splashing" back up
Perfusion (artificial perfusion)	C4-F#4-C4 "PER-FU-SION"	C4-F#4-C4—C4-F#4 (repeated) "PER-FU-SION A-LARM"	Artificial sound; tri-tone; similar to "yo-ee-oh" of the Munchkins in "The Wizard of Oz"
Power failure	C5-C4-C4 "POW-ER FAIL" "GO-ING DOWN"	C5-C4-C4—C5-C4 (repeated) "POW-ER GO-ING DOWN"	Falling pitch as when the power has run down on an old Victrola

4.2.2 Visual Urgency

The visual modality can depict urgency cues through various techniques. Research has indicated that the most effective visual techniques to attract attention in urgent situations can be conveyed through messages with movement (e.g. blinking/flashing, position change), size and shape differentiation, texture and brightness (Chung & Byrne, 1997; Ho et al., 2001). **It is important to note that these techniques should be used sparingly, since they have very strong attention capture effects.**

4.2.3 Tactile Urgency

Brewster and Brown (2005) used different levels of roughness of vibrotactile stimuli to encode the priority levels of alerts in *tactons*. The roughness was created by using amplitude modulate sinusoids, with increased roughness being caused by decreases in frequency. Therefore, the same method of implementation through vibrotactile displays can be used to present various levels of urgencies. Van Erp and Self (2008) claim that the density of tactors in an area can be used to indicate the priority of a message. A small number of tactors located at a specific area of the skin can be activated to present a low priority threat, while activation of a large number of tactors located spatially close to each other can be used to indicate a high priority threat.

Different levels of intensity or amplitude of vibration can also be utilized to present the values of variables. For example, proximity of aircraft in a restricted area can be represented through different amplitude levels of vibration (Jones & Sarter, 2008). Therefore, **it is also possible to encode different levels of urgency in the form of different intensity or amplitude levels of vibrations**. Van Erp and Self (2008) state that research has suggested that various frequency and amplitudes can also convey target information for pilot operators. For example, spatial distances from targets or priority of targets are possible pieces of information that can be conveyed through this dimension of tactile displays. It is important to note that there are uncertainties about the perception of change in frequency by participants. Therefore, it is best to use the frequency at a fixed level.

4.3 Concluding Remarks

Auditory information has been widely used by interface designers, even before research into multimodal displays began in earnest. In this section, we have shown that one of the research questions that is at the forefront of auditory display research is how individuals make sense of the auditory stimuli that they perceive. We have described how auditory signals can be coded according to two different continuums: symbolic-analogic and complexity. Codings that make use of analogous connections between the physical signal and the input data tend to be more memorable, and may also result in faster processing times. However, symbolic codings are more flexible and may be used to represent concepts that do not have adequate real-world counterparts.

There has been strong evidence that there are intuitive mappings that exist between different types of data and their auditory representations. This is especially evident in the presentation of urgency information where some auditory characteristics, such as intensity and speed have found to be very indicative of perceived urgency. Overall, it is possible to portray urgency messages through the auditory modality as well as the visual and tactile modalities. Each modality has demonstrated that they have different constraints in presenting urgency information. For example, it appears that participants can distinguish differentiations in frequency levels in the auditory modality more readily than in the tactile modality. However, besides the Sanderson et al. (2006) paper, there are few studies comparing perceived levels of urgency across different modalities. While there is evidence that some amodal characteristics can be learned and transferred across modalities, further research into this topic would provide valuable insights when designing to support proper attention direction in a multimodal interface.

5 Crossmodal Attention

In the introduction to *The Multisensory Driver*, Ho and Spence (2008) describe the importance of studying crossmodal attention when designing interfaces for vehicles,

> *Humans are inherently limited capacity creatures; that is, they are able to process only a small amount of sensory information that is typically available at any given time... The limited capacity of spatial attention to process sensory information in humans raises important constraints on the design and utilization of, for instance, vehicular information systems...The ability of drivers to attend selectively and their limited ability to divide their attention amongst all of the competing sensory inputs have a number of important consequences for driver performance. This, in turn, links inevitably to the topic of vehicular accidents.* (p. 1)

In the context of designing interfaces for UAV GCSs, a limit in the ability of the operator to attend to required stimuli could lead to missed mission objects, or a loss of vehicle. In the following sections, we present literature related to how attention is directed in each modality, and between different modalities to support the idea of attention mapping which was suggested in the EID section.

- Section 5.1. Presents four general theories which explain how multimodal sensory events are handled and how attentional resources are allocated. They include: the theory of independent modality-specific attentional resources, the theory of single supramodal attentional resources, the theory of separable but linked attentional systems, and the theory of hierarchical supramodal plus modality specific attentional systems. In addition, several Bayesian models for predicting attention division in multisensory events are presented.

- Section 5.2. Addresses the issue of cue conflict situations and how humans respond to such events. The section separates the areas of research into warning signals which are primarily governed by *exogenous attention*, and monitoring tasks which are primarily governed by *endogenous attention*. From an interface design perspective, a review is provided which offers guidelines for the placement of tactile stimuli, the combination of sensory modalities for maximum effectiveness, the use of multimodal cues for focused versus divided attention tasks, and the effect of sensory bias on conflict resolution.

- Section 5.3. Addresses the issue of an operator's ability to attend to multiple channels of information. This includes the effects of *load stress* and *speed stress*, as well a discussion of complacency in highly reliable sources.

- Section 5.4. Addresses pre-attentive processes that interfaces can exploit to reduce the attentional load in multimodal interfaces.

- Section 5.5. Provides concluding remarks.

5.1 Crossmodal Attention Resource Models

A common aspect of both recent and past research on crossmodal attention is the concept that resources can be combined and allocated according to different theories of attention. Within the literature, there are four commonly cited theories of crossmodal attention. These theories are the division of resources based on the concept that each modality is governed by an independent process, the single supramodal attention system, the independent plus linked attentional systems, and the hierarchical supramodal plus independent attentional systems (Spence, 2009). The four models for attention can be seen in the figure below. The following sections present each theory in detail. Following this we discuss Bayesian models, which present a statistical approach for modeling attention resources.

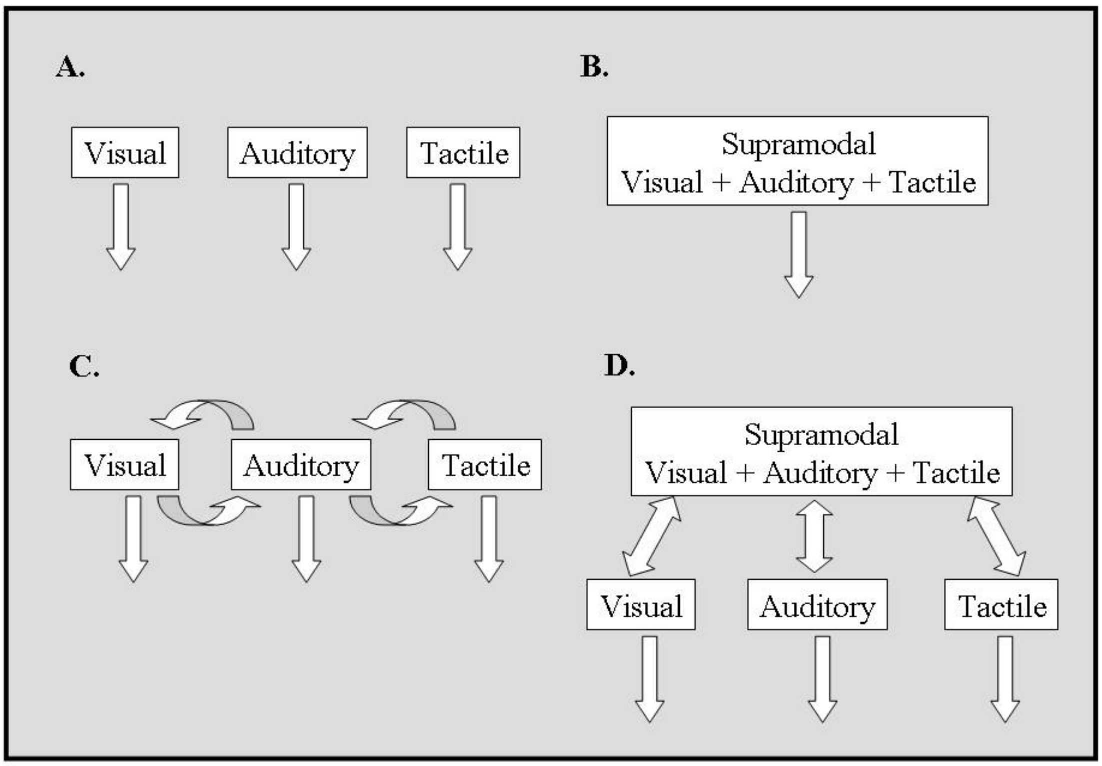


Figure 23: Models for Crossmodal Attention (Spence, 2009)

5.1.1 Independent Modality-Specific Attentional Resources

As described by Sarter (2007), the theory of independent modality-specific attentional resources suggests that there are separate fixed-capacity resources for information processing. Thus, the visual, auditory, and tactile attentional system are relatively independent. *Multiple resource theory* (MRT), introduced by Wickens (1984), encompasses the independent modality-specific attentional resource theory. The main premise of MRT is that humans do not have a single source capable of information processing, but a number of resources that can be accessed concurrently. Wickens suggested that low performance is characterized by a lack of available pools of resources, indicating that cognitive resources are limited. However, he also explains that the connection between workload and performance is complicated. For example, **high workload can cause performance to decrease, but low workload can cause complacency** (Moray, 1981).

Experimental support for independent modality-specific attentional resources has been found. For example, Rees, Frith, and Lavie (2001) investigated whether load effects are governed by modality-specific attentional resources, or whether they are caused by visual and auditory interactions. Two observations were made: that both an activation of the visual cortex and a robust motion after-effect is caused by unrelated dynamic visual stimuli during a simultaneous auditory stimuli presentation. These observations were made for both high and low auditory load conditions. More recently, Brill, Mouloua, Gilson, and Rinalducci (2008) used a multimodal secondary loading task paradigm to examine whether each modality drew from a separate pool of reserve cognitive capacity. Participants were required to complete a visual monitoring test while also concurrently completing a secondary task involving signals in different modalities. They found that performance and subjective workload suffered for the visual secondary task when compared to secondary tasks in the tactile and auditory modalities. Taken together, these findings agree with the theory that **the main source of attentional restrictions is caused by modality-specific subsystems**.

5.1.2 Single Supramodal Attention Systems

In contrast to the MRT and models that involve independent modality-specific attentional resources, some researchers had favoured models which claim that there is only a single system that controls attention, and consequently a single pool of resource that is shared amongst all modalities. The theory of a supramodal attention system was first evaluated in a ground-breaking study by Farah et al. (1989). The concept of a supramodal attentional system includes the idea that humans can attend to only a single location at any time, and are not able to divide their attention to different location simultaneously. However, this focused attention location may be shared across different sensory modalities (Santangelo, Fagioli, & Macaluso, 2010).

In Farah et al.'s study, the authors investigated whether spatial attention was separated into a modality-specific subsystem, or if a supramodal spatial attention system exists, as described above (Farrah et al., 1989). An experiment was derived to address this question; participants suffering from parietal lobe lesions were required to respond to visual stimuli, which were preceded by either an auditory or visual cue. The parietal lobe in the brain is responsible for integrating sensory information from different sensory modalities. All cues were presented on the side of body without the lesion and were non-predictive (50% chance of being correct). In both cue situations, participants were slower to respond to invalidly (incorrectly) cued targets which occurred on the side of the body opposite of the lesion. This observation indicated that there was attentional disengagement impairment for visual targets with auditory cues in addition to the expected disengagement impairment for visual targets with visual cues. Thus, it can be concluded that the parietal lobe's attentional mechanism is based on a representation of space where both visual and auditory stimuli are represented. This result is evidence against modality-specific attentional resources, and supports the above theory of the existence of a supramodal attention system (Farah et al., 1989).

More recent work has suggested that common cerebral regions may promote the construction of higher order representations in working memory for both visual and tactile information. These findings support the theory of supramodal organization in memory applications (Gallace & Spence, 2009).

5.1.3 Separable but Linked Attentional Systems

Spence and Driver (1997) suggested that attention controls for different sensory modalities are connected, but are also capable of acting independently. This theory attempts to address discrepancies seen in the earlier models where strong links between different modalities had been shown, but it also provides evidence for the ability to direct attention to different spatial locations for different modalities (Spence & Driver, 1996).

To investigate the existence of crossmodal links, Spence and Driver (1996) completed a series of experiments which studied the connections in endogenous (goal directed) spatial orienting in hearing and vision. In particular, Spence and Driver were interested in studying covert orientation of attention, where reorientation of the body, head, or eyes was not required. The participants were required to respond to auditory and visual stimuli with elevation guesses (either up or down). There were several important observations in this study. Firstly, when participants were aware that the stimuli would be located on a specified side of the body, response times were shorter, regardless of the modality of the target. Secondly, when participants were aware of the modality of the target, a shift of attention occurred in the other modality, which also resulted in shorter reaction times. Lastly, when participants were aware that the targets would be presented in two modalities, the auditory and visual attention was often divided. **These observations support the hypothesis that endogenous covert spatial attention does not occur solely within a supramodal system, and neither do the modalities act independently. Rather, the authors suggest that there are strong special links between visual, auditory, and tactile attention (Ho & Spence, 2008)**.

5.1.4 Hierarchical Supramodal Plus Modality Specific Attentional Systems

Lastly, a hybrid model has been proposed which encompasses the interconnections of the modality-specific attentional resources and the attention systems of the supramodal modal. The work of Posner, Spence, and Driver (1996) suggested that a supramodal plus modality-specific attentional system may also describe their own experimental observations. They describe this model as one where the unimodal attentional subsystems supply into a higher-level supramodal system. Therefore, individual modalities may have their individual pools of resources which are used when tasks are modality specific, while tasks that require crossmodal attention may draw from a supramodal pool of attentional resources. One application of this work was in describing the *Colativa effect*, an example of visual dominance that will be described in more detail later in this section.

5.1.5 Bayesian Modeling Approaches

As listed previously, there are presently four conceptual models for describing the division of attentional resources. However, there is presently no general theory for describing the mechanisms of attention, and more specifically, how conflicts between competition and conflicting cues in different modalities are resolved (Beierholm, Kording, Shams, & Ma, 2007). However, past research has indicated that the determination of the spatial properties of multisensory stimuli indicate that people integrate multimodal inputs using a statistically optimal method, which includes a weighting system for each sensory input (Ley, Haggard, & Yarrow, 2009). Thus, contemporary research has begun to focus on the determination of statistical models for decision resolution from several multisensory inputs using a Bayesian modeling approach (Beierholm et al., 2007). Bayesian models make use of prior knowledge to predict the probability of a future event occurring. Some of the models presented in recent research are presented next, along with a comparison of each model.

5.1.5.1 Maximum-Likelihood Estimation

In one of the fundamental paradigms in contemporary Bayesian modeling, it is assumed that a common source exists for the incoming multisensory stimuli (Beierholm et al., 2007). The strategy of the model is to introduce a small conflict between multisensory cues, which allows an estimate of the effect of the common source stimulus to be estimated from both the common source and the small discrepancy. The estimate of the effect of the common source can be determined from the knowledge that the percept deduced from the integration of different sensory cues will lie somewhere between the precepts deducted from each cue individually. The assumption is that a higher weighting will be placed on the most reliable cue, and thus the representation of what is perceived by the individual will be closest to the representation that is obtained from the most reliable cue (Ma & Pouget, 2008).

5.1.5.2 Cue Integration with Consideration of Prior Knowledge

Roach, Heron, and McGraw (2006) investigated the effectiveness of the concept of the maximum-likelihood estimation and found that the model was not consistent with their results. In the study, participants were asked to respond to stimuli in one modality while ignoring conflicting rate information in another modality. The authors found that there was a slow transition from partial cue integration to complete cue isolation, which was not consistent with the maximum-likelihood estimation model.

Thus, a revised model was designed to consider the prior knowledge regarding the connections between multisensory signals when determining the degree of integration, taking into account the predictiveness or non-predictiveness of information in different modalities as a priori information.

Thus, a strategy is determined for balancing the benefits accumulated from sensory estimates determined from a common source, as compared with the costs of combining information caused by independent objects or events (Beierholm et al., 2007; Roach, Heron & McGraw, 2006).

5.1.5.3 Causal Inference Model

For the causal inference model, the assumption that two different sensory signals are caused by the same source is invalid. In this model, there can be one or two sources, and the number of sources is considered a parameter which can be inferred from the cues presented (Beierholm et al., 2007).

The model allows the observer to consider two hypotheses about the multisensory event: that they have a common cause or that they have separate, independent causes. The Bayesian model considers that the observer computes the probability of each hypothesis which is dependent on the noisy sensory signals of the trial and the prior information about the presence of a common cause (Ma & Pouget, 2008).

5.1.5.4 Comparison of Models

To investigate the effectiveness of each model, Kording et al. (2007) completed a psychophysics experiment where participants were required to respond to a short visual and auditory stimulus. Participants were required to indicate the perceived position of visual and auditory stimuli. Kording et al. found that the causal inference model fit the human data better than the other models proposed.

It should be noted that other studies have also shown that the Bayesian model for cue integration with the consideration of prior knowledge also fit human data quite accurately (Ma & Pouget, 2008). Although Bayesian models are improving in their ability to correctly model human perceptual experience, the accuracy of current Bayesian models is still inadequate to predict complex human behaviour.. Thus, improvement is required in this area.

Bayesian modeling is applicable to interface design because it allows designers to predict the effectiveness of multisensory inputs computationally instead of experimentally. Thus, theories on multisensory input combinations can be tested without a large time and cost investment. Also, having a model for how operators will integrate and perceive information across different modalities is very important for the design of multimodal interfaces. If an interface was designed to provide redundant information across modalities, then the operator's perception of the information will be based on this "integrated" data source. The modality of presentation and the operator's a priori knowledge of how information is presented within each modality can be modeled using Bayesian approaches to predict how they will perceive the information.

5.2 Conflicting Cue Situations

When designing a multimodal interface, it is important to consider how the operator will respond to stimuli overload and conflicting information from different sensory modalities. The goal is to portray information as clearly as possible; however, cue conflicts can occur where two different sources may provide contradictory or inconsistent information. For example, a situation may occur where information from the visual modality conflicts with information provided by the tactile modality, either at a perceptual or semantic level. How do humans handle this conflict of information?

In interface design, there are two modes of attention which can be leveraged by interface designers to guide operators to the most relevant information in an interface. These modes of attention are associated with two types of tasks that operators often are required to accomplish. The first task is one where operators must respond to unexpected events, such as warnings. This situation is primarily governed by *exogenous attention*, which refers to attention being drawn without conscious attention. The second task is where an operator is expected to continually monitor values or states, such as in a supervisory control situation. This task is primarily governed by *endogenous attention*, which refers to the voluntary control of attention.

5.2.1 Exogenous Attention: Responding to Unexpected Events

Often in operational applications, a situation occurs where the operator’s attention is directed to a fault or warning, while the operator is monitoring something else. This situation is characterized by exogenous attention. Exogenous orienting is described as the stimulus-driven, or bottom-up, directing of a person’s attention where the reflexive orienting of attention occurs as a result of external stimulation (Ho & Spence, 2008). Note that it is possible to present warning signals in an endogenous manner, where attention is directed by cueing the area of focus.

Exogenous attention is governed by stimulus-driven attentional control, which is associated with the response to perceptual characteristics of the stimuli instead of the semantic meaning of the stimuli. These direct cues are associated with stimuli that occur directly at or in the vicinity of a potential target location. Therefore, warning events tend to be cued by stimuli which are processed quickly based on highly salient characteristics. This type of attention control, compared with goal-driven attentional control, is much faster to perceive. **The effectiveness of stimulus-driven attentional control is at its maximum – approximately 100 ms after the warning event occurs (Wright & Ward, 1954 as cited in Wright & Ward, 2008).**

5.2.1.1 Placement, Timing, and Loading of Stimuli for Maximum Effectiveness

When designing a multimodal interface, the location and modality of each stimulus must be carefully chosen as it affects the ability for the stimuli to capture the user’s attention in addition to other factors such as task. Recent research has shown that warning signals placed very close to or on the body of an operator are more effective than stimuli placed in the extrapersonal space. This is because the brain treats stimuli in this region as being more behaviourally relevant and demanding of immediate attention as compared with stimuli in the extrapersonal space (Previc, 2000 as cited in Ho & Spence, 2009). This can be related to the primitive *"margin of safety"* which exists around the body for defensive purposes (Ho & Spence, 2009). This margin of safety is widely referred to a *peripersonal space*, which is defined as the space immediately surrounding the body. In experiment 2 of a study by Ho and Spence (2009), participants were asked to respond to different types of auditory, tactile, and visual stimuli. The origin of the different stimuli was varied to investigate the effect of stimuli placement on response time. Ho and Spence (2009) found that auditory stimuli that originated from locations close to the participant resulted in the fastest response times when compared to distant auditory stimuli. The auditory stimuli were also more effective at reducing response time than the tactile and visual alerts. However, tactile alerts out-performed visual alerts.

This provides evidence that **the use of vibrotactile and auditory warning signals can improve an operator's response to faults** (Ho & Spence, 2009). Other recent studies have also compared the ability of alerts in different modalities (e.g. auditory, tactile, and combinations of different modalities) to draw the operator's attention. For example, Scott and Gray (2008) investigated the effectiveness of rear-end collision warnings which were presented to different sensory modalities as a function of warning time. The participants were asked to respond to four warning conditions: no warning, visual, auditory, and tactile. The warnings were activated when the time-to-collision reached a value of three or five seconds. Subsequently, the driver's response time was measured by using the amount of time elapsed until brake initiation. The study found that of the four conditions, tactile warnings were the most effective in prompting participants to respond to potential rear-end collision events. Also, tactile warnings elicited the shortest response times (Scott & Gray, 2008).

However, it should be noted that vibrotactile and auditory stimuli are not effective in every situation. For example, the exogenous capture of attention (stimuli-driven) can be dominated by endogenous control (goal-directed) in some situations. **With spatially non-predictive visual, auditory, and tactile cues, the effect of the stimuli becomes ineffective when participants are given a secondary, attention-demanding perceptual task concurrently. The likelihood that any spatial cue will capture a person's attention depends on its salience relative to the current focus of attention** (Spence & Santangelo, 2009). Therefore, there is some evidence that the effect of exogenous capture of attention via unimodal stimuli is decreased with increased workload. **However, bimodal stimuli that originated from the same location continue to capture attention in high perceptual workload conditions** (Spence & Santangelo, 2009). This suggests that bimodal/multimodal stimuli may be better at alerting operators in high workload conditions. It is important to note that Spence and Santangelo (2009) also state that multimodal cues did not outperform (response time and accuracy) unimodal cues in low workload conditions.

5.2.1.2 Redundant Warnings: Modality Choice and Conflict Resolution in Exogenous Attention

As mentioned previously, it has been shown that warning signals are more effective when presented in the peripersonal space (close to or on the body). Current research indicates that different types of warning signals can be combined and configured to obtain effective signals for alerting the user (Spence & Ho, 2008). However, some current research also indicates that detrimental effects can occur if modalities are combined incorrectly. For example, work completed by Kitagawa, Zampini, and Spence (2005) showed that auditory distracters can interfere with tactile left and right discrimination when the auditory stimuli are placed close to the head. However, Kitagawa et al. (2005) found that the effect does not occur when auditory stimuli are placed further from the head. Thus, special consideration needs to be taken when considering the use of redundancy over modalities and the placement of the sensory stimuli.

Also, it has been demonstrated that warning signals placed in the same spatial location can aid in attracting the attention to a stimulus in another sensory modality (Kitigawa, 2006; McDonald, Teder-sälejärvi & Hillyard, 2000; Ho, Santangelo & Spence, 2009). McDonald et al. (2000) completed research that showed that a sudden sound in the same location as a sudden flash improves the number of successful detection of a visual stimulus. The researchers utilized signal detection measures, as opposed to reaction times, to investigate the perceptual or post-perceptual processing of the nearby visual stimulus.

The above examples provide results that are both different and conflicting; some research states that redundant warnings are detrimental to performance while other research states the opposite. In much of the past research, there has been some confusion pertaining to exogenous crossmodal shifts and modality-specific properties of the systems involved in the encoding of spatial locations. There are presently several hypotheses as a result of current research (see 5.1 for description of models). Unfortunately, more research needs to be completed to determine parameters for when sensory combinations are beneficial. For interface designers, Bayesian modeling (see Section 5.1.5) provides a promising method for estimating the effects of multisensory integration.

5.2.2 Endogenous Attention: Monitoring of Continual Variables

For operators, it is often required or advantageous to monitor values and states continually, in addition to responding to unexpected warning events. In this situation, the operators voluntarily choose where to focus their attention. Endogenous orienting includes the voluntary shifting of a person’s attention which is driven internally by top-down control (Ho & Spence, 2008).

An example of endogenous attention occurs when a person is instructed to focus their attention on a defined target (Pattyn, Neyt, Henderickx, & Soetens, 2008), such as an operator being required to monitor the altitude of a UAV. This mechanism is characterized by a *controlled* processing mode, because the focus of attention is determined by the person’s goals and expectancies (Pattyn et al., 2008).

Endogenous attention is governed by goal-driven attentional control, which is associated with the response to symbolic cues. These symbolic cues are associated with stimuli that indirectly point to a potential target location. Thus, some processing of the symbolic cue must first be done to understand the meaning of the cue. Once this cue is understood, the attentional system is then directed towards the indicated location. The response of this type of attention control is much slower than stimulus-driven control. This is because the **effectiveness peaks at 300 ms – nearly**

100ms later than the approximate effectiveness time of stimulus-driven attentional control (Wright & Ward, 1954 as cited in Wright & Ward, 2008).

5.2.2.1 Redundant Warnings: Modality Choice and Conflict Resolution in Endogenous Attention

In general, there are two approaches in research which address the task of human supervisory control. The first approach is to study the attentional mechanisms which process signals that come from the location or sensory modality that the operator is currently focused on. However, in this approach, information located outside the field of attention is ignored. The second approach addresses the mechanisms used to select and react to the information which are located in separate spatial locations (Santangelo et al., 2010).

The first approach addresses the possibility that multisensory cues may be used to orient a human's attention to one location. When this occurs, the different sensory modalities will be used to orient the participant's attention to the same spatial location. Wright and Ward (1954 as cited in Wright & Ward, 2008) suggested that a cue in one modality would cause a label, or a tag, to be associated with the spatial location that the cue occurred within the multimodal spatial map. However, when multiple cues in different modalities are used to direct attention to the same location, the first label would cause the reorientation of attention to be inhibited leading to slower response times (Wright & Ward, 1954 as cited in Wright & Ward, 2008). This suggests that some interference can occur when redundant multisensory cues are used to orient attention.

In the past, monitoring of multiple modalities has been largely focused on attending to different senses at one spatial location, such as having auditory and visual stimuli produced from the location of a desktop computer. However, Santangelo et al. (2010) investigated whether monitoring and processing different sensory modalities is more efficient when attention is spatially divided than when focused at a single location. It was found that in-parallel processing is more effective for spatially divided stimuli in different sensory modalities, which is an important concept in the design of multimodal interfaces. In the study, participants were asked to simultaneously monitor vision and audition in two cases: focused attention, and divided attention. An additional case where a single modality was monitored at one or two locations was also used for comparison. The study showed that the cost of monitoring two modalities versus one modality decreases when spatial attention is divided between two separate locations compared with focused attention. In addition, neuroimaging data showed that when participants' monitored two modalities at different location, there was an increased activity in the posterior-parietal cortex. Activation in the posterior-parietal cortex had also been found in other studies of spatial attention for both visual and auditory stimuli (Santangelo et al., 2010). However, there was no specific brain region utilized when participants' were involved in the focused attention situation. From these results, it was concluded that the engagement of the posterior-parietal cortex and the stronger use of the modality-specific resources allow for effective in-parallel processing when attention is spatially divided (Santangelo et al., 2010). The authors suggest that the role of the posterior-parietal cortex may be to coordinate multiple modality-specific attentional resources.

In regards to multimodal interface design, the above research will aid designers in deciding which applications to use multimodal cueing. **The work by Santangelo et al. (2010) shows that multimodal cueing is advantageous for two separate spatial locations and the work of Wright and Ward (1954 as cited in Wright & Ward, 2008) suggests that in some situations, multimodal cueing can disadvantageous for different modalities in the same spatial location due to inhibition effects .**

5.2.3 Interaction between Endogenous and Exogenous: Decision Conflict and Attention Issues

Since the 1960s, research regarding intersensory conflict resolution has been performed in two main areas: the ability of a human to adapt to multisensory conflict over time through adaptation, and the immediate response that humans have to multisensory discrepancies (Welch & Warren, 1980).

Past studies have indicated that one or more modalities tend to bias the others in multisensory conflict situations (Beierholm et al., 2007; Helbig & Ernst, 2007). This is also referred to as crossmodal bias, which occurs when a person localizes an input based on one modality, but ignores the input from another modality (Vroomen, Bertelson, & de Gelder, 2001). An example of this was shown by Lederman, Throne, and Jones (1986) in an experiment on the effectiveness of the visual and tactile modalities with regards to determining the spatial density and roughness of a textured surface. In the study, participants were asked to determine the spatial density and roughness of a textured surface using either touch, vision, or a combination of both touch and vision. The results showed that there was a strong influence from the visual modality for determining the spatial density property, but for the surface roughness property, the tactile modality was more influential (Lederman et al., 1986).

The Lederman et al. (1986) study, as well as other research, shows that intersensory bias exists (Beierholm et al., 2007; Spence & Ho, 2008; Vroomen et al., 2001; Welch & Warren, 1980); however, what determines which modality dominates?

First and foremost, humans often tend to process information more readily in the visual modality. Vision, also has a high bandwidth of information transfer, which leads interface and display designers to often overload the visual modality with information (Hager, Kriegman, & Morse, 1998; Hameed & Sarter, 2009). According to Lukas, Philipp, and Koch (2010), visual dominance is the tendency in which people prefer to direct their attention towards the visual modality. Colavita (1974) conducted various experiments that suggested humans have a visual sensory dominance and discovered the tendency of humans responding more often to visual stimuli compared to auditory stimuli during speeded discrimination tasks; this phenomenon was later named the Colavita effect. The Colavita effect is the tendency to respond to visual targets/stimuli over other modality targets (Colavita, 1974).

According to Koppen and Spence (2007a), there are various variables that have been proven to modulate the Colavita effect including stimulus probability, spatial coincidence, and audiovisual asynchrony. For example, experiments manipulating stimulus probability significantly decreased the magnitude of the Colavita effect. This finding is consistent with literature on attention which states that increasing the frequency of specific targets (e.g. bimodal targets) will result in directing participants' endogenous attention towards that specific target, improving performance in speeded discrimination response tasks (Koppen & Spence, 2007a). Based on many studies supporting the concept of visual dominance, Lukas et al. (2010) suggested a possible explanation for the visual dominance effect in terms of attention direction. This explanation states that visual stimuli are not as salient as other modalities. Therefore, humans focus their attention more readily on visual stimuli to compensate. It has been established that attention does in fact play a role in the visual dominance effect; however, the impact of visual dominance was found to be influenced through attention manipulations. For example, when participants' attention were directed to auditory stimuli by increasing the auditory stimuli's probability and proportion, the Colavita effect was not as apparent (Sinnet, Spence & Soto-Faraco, 2007 as cited in Lukas et al., 2010). Another modulating factor is spatial coincidence where the Colavita visual dominance effect was reduced when auditory and visual stimuli were presented from different positions compared to when they were presented in the same position (Koppen & Spence, 2007c). Furthermore, evidence also indicates that the Colavita visual dominance effect can be modulated by audiovisual asynchrony. This study found that the Colavita's effect was affected by the temporal order in which the visual and auditory stimuli were presented in the bimodal targets. It was found that the Colavita effect was larger when the visual stimuli was presented first compared to when the auditory stimuli was presented first (Koppen & Spence, 2007b). **Multiple studies have shown how various factors can result in the visual dominance effect being attenuated and vary performance in terms of error rates and reaction times across vision and audition**. **Thus, multimodal interface designers should consider these results when attempting to take advantage of visual dominance effects and possible variables that can reduce the magnitude of visual dominance**.

With regards to when and what sensory modality will dominate, studies have demonstrated that the dominant sense is determined by the situation and the properties evaluated. For example, **the determination of properties such as size, shape and spatial location, referred to as macrospatial tasks, are dominated by the visual modality. In contrast, microspatial tasks can be dominated by the auditory modality in temporal tasks that require one to determine the rate of duration. In addition, there are situations where sensory dominance is not always clear. For instance, in tasks that require the determination of surface properties, intersensory bias can change depending on the texture parameter (i.e. spatial density versus roughness) investigated (Lederman et al., 1986).**

5.3 Differentiating Between Two or More Channels of Information

The previous section addressed the issue of conflicting sensory information, which can be reduced and alleviated by proper interface design decisions, including the correct placement of stimuli and the intelligent integration of sensory information. The issue addressed was *how* to present information. However, interface designers must also consider the concept of *how much* information to present for maximum effectiveness.

Current research in time-sharing (performing more than one task simultaneously) applications indicated that there is no strategy that can guarantee the timely detection of all faults if attention must be shared amongst two or more channels. Past research investigated the maximum number of signals that can be detected, but there is no assurance that all will be detected (Moray & Inagaki, 2000).

5.3.1 Selective Attention, and the Effects of Speed Stress and Load Stress

Selective attention filters out unnecessary and irrelevant information and processes only sensory information that is relevant to the observer (Huffman, 2007). As the number of channels of information increases, performance will decline, even if the signal rate remains constant. Two types of stress, load stress and speed stress, negatively affects a participant's performance in these situations (Goldstein & Doftman, 1978). *Load stress* is the stress caused by increasing the number of channels over which is information is presented (Gawron, 2008). *Speed stress* is the stress caused by changing the rate of signal presentation (Sanders & McCormick, 1993).

Goldstein and Dorfman (1978) investigated the effects of speed stress and load stress, and found that an increase in both types of stresses led to significantly poorer performance. In the study, participants were asked to respond to moving visual stimuli that entered critical zones over three visual displays. It was found that when participants were required to interact with one display, which represents a **low load stress condition, increases in speed stress did not have a strong effect on performance. However, as the load stress increased to the use of two or three displays, increases in speed stress had a significant effect on reducing performance**.

In addition to the effects of the types of stresses, investigations have been performed into which channels of information dominate the participant's attention. When humans are required to sample multiple channels of information, the attention tends to be focused at signals which occur more frequently (Sanders & McCormick, 1993). Due to the limitations of human memory, it is common for participants to forget to sample a channel when multiple sources are present (Moray, 1981). Also, humans tend to sample a channel more when they remember the previously displayed value of the source when it was previously sampled (Sanders & McCormick, 1993).

5.3.2 Complacency

Also, similarly to the handling of conflict situations, often the brain implements statistical algorithms to make decisions. This is also true in the situation of multiple channel handling, where the brain must make decisions on which channel to sample, and how often to sample each channel of information. Moray (1981) has suggested that what many researchers deem as complacency can be attributed to a rational strategy. When a participant is required to attend to multiple channels of information simultaneously, they may avoid sampling more reliable sources (e.g. sources with much less variability or sources that have higher event rates) with the goal of reducing workload and attending to more volatile problems/sources. However, research has shown that operators do not sample efficiently even when the underlying probabilities of encountering a fault within a source is known (see Wickens & Hollands, 2000 for a review). Nearly twenty years later, Moray and Inagaki (2000) revisited this issue, realizing that in real systems, one is not content with this type of behaviour by operators. Although the decision making should be rational, it should be rationally skeptical; meaning that a source should never be trusted, even if it has never failed. From this, the question arose: at what frequency should a 100% reliable source be sampled? Moray and Inagaki suggested that one approach would be to model the source, both causally and mathematically. The model would account for a worst case situation where the operator would be required to intervene and take action towards preventing the fault from becoming a disaster. They proposed a possible model where a system which has never encountered a fault should be sampled at a frequency,

$$f \cong (T - t)w \tag{3}$$

Where T is the time from the occurrence of a fault until the dangerous consequences are unavoidable (the incident is unrecoverable), t is the time required to take action to prevent the unrecoverable consequences, and w is a weight related to the severity of the consequences of an unrecoverable accident. This model can be used as a guide for how an ideal operator should sample a visual display.

In regards to the design of multimodal interfaces, the existence of a timing scheme for checking reliable sources may be used to integrate exogenous and endogenous orientation in warning signals. For example, an operator involved in monitoring a visual interface for signal A, and a message may come up reminding the operator to check the status of signal B for every certain time period. This would ensure that both channels would be monitored. In addition, the knowledge of load and speed stress can assist designers in deciding the optimal amount and speed of information to transfer to the operator simultaneously.

5.4 Pre-Attentive Characteristics of Different Modalities

The concept of *attentional mapping*, as described by Sanderson et al. (2000) and reviewed previously in the EID section, requires that interface designers direct attention to the most relevant pieces of information when it is required. Interface designers must also reduce the distractibility of data that is not needed at a given time. This is especially true for information presented in the auditory and tactile modalities since these channels cannot be “turned off”. To this end, a strong understanding of pre-attentive perceptual processing, and attention capture is needed. Healey, Booth, and Enns (1996) describe pre-attentive visual processing as “cognitive operations that can be performed prior to focusing attention on any particular region of an image”. In the visual search literature, pre-attentive processing is regarded as a parallel process that has an unlimited capacity. Given the descriptions of crossmodal attention described earlier in this section, the possibility of exploiting an unlimited capacity channel for interface design is desirable. The amount of attentional resource that can be deployed is limited even for the independent modality-specific attentional resource model. In the visual field, detection of a “featurally defined stimulus” (e.g. a blue target amongst red distractors) based on its defining feature (colour) occurs even when attention is directed elsewhere (Smith & Ratcliff, 2009). Thus, individuals can gain information from stimuli without needing to direct attentional resources towards the source. Therefore, pre-attentive features across modalities can be used to communicate information in a multimodal interface without using the operator's limited attentional resources.

Currently, there is very limited literature on pre-attentive processes in the tactile modality, and many of the tactile coding principles described earlier were tested with participants directing focal attention to the tactile modality. However, we are able to draw from knowledge of pre-attentive processes for the visual and auditory modalities to possibly gain some insights that can be applied to the tactile display design. In visual displays, it is essential to present information in a way that users are able to absorb meaningful data with minimal effort. This goal can be attained through the utilization of pre-attentive visual features. Healey et al. (1996) presented a chart depicting visual features that have been utilized to perform pre-attentive tasks. **These pre-attentive features should be considered when designing visual displays since they can be used to communicate information without the need for focal attention**. However, these pre-attentive features may not always be applicable, and are largely dependent on the context of the display.

Table 7: Visual Pre-Attentive Features (Adapted from Healey et al., 1996)

Feature	Author
line (blob) orientation	Julész & Bergen [1983]; Wolfe [1992]
Length	Triesman & Gormican [1988]
Width	Julész [1985]
Size	Triesman & Gelade [1980]
Curvature	Triesman & Gormican [1988]
Number	Julész [1985]; Trick & Pylyshyn [1994]
Terminators	Julész & Bergen [1983]
Intersection	Julész & Bergen [1983]
Closure	Enns [1986]; Triesman & Souther [1985]
colour [hue]	Triesman & Gormican [1988]; Nagy & Sanchez [1990]; D'Zmura [1991]
Intensity	Beck et al. [1983]; Triesman & Gormican [1988]
Flicker	Julész [1971]
direction of motion	Nakayama & Silverman [1986]; Driver & McLeod [1992]
binocular lustre	Wolfe & Franzel [1988]
stereoscopic depth	Nakayama & Silverman [1986]
3-D depth cues	Enns [1990]
lighting direction	Enns [1990]

In terms of pre-attentive processing in the auditory modality, research has suggested that *sonification* is a candidate for pre-attentive processing. Continuous signals, such as those provided in the sonification, eventually fade out of focal attention which is then monitored pre-attentively, requiring no use of attentional resources. However, when a continuous signal experiences a deviation, the operator will orient their attention toward the audio signal (Spain &

Bliss, 2008), facilitating the transfer of the sonification back into focal awareness. Possible auditory deviations that can direct an individual's attention include pitch, duration, pulse, and tempo (Spain & Bliss, 2008). These parameters can be used as a tool for interface designers when they are required to support operators who must manage multiple tasks concurrently. The lack of research on pre-attentive processes in the tactile domain can be addressed by using techniques in the visual and auditory domain. Pre-attentive processes are a useful human capability that can be used to help operators use their attentional resources sparingly. However, the issue of capturing attention, which is required by attentional mapping, still needs to be addressed in greater detail.

Humans often encounter countless cluttered visual scenes with various objects that compete to be noticed by the observer. Attention is the underlying mechanism that is used to direct humans to relevant information. In order to save time and effectively process information, selective attention must be applied to allocate their attention to the relevant and useful visual cues within the scene. As stated before, selective attention filters out unnecessary and irrelevant information and processes only sensory information that is relevant to the observer (Huffman, 2007). When discussing attention, we must consider factors such as different stimuli characteristics that are capable of capturing attention. Interface designers can make use of these characteristics to ensure that the operator's attention is shifted to the correct modality and spatial location when the data is relevant to the operator (as defined by the attentional mapping).

In vision, the most eminent theory explaining what types of stimuli captures attention is the new-object hypothesis which states that the only type of stimuli that can automatically capture one's attention is when a new visual object is presented in the visual scene. According to the new-object hypothesis, when an individual scans a visual scene, visual objects are indexed and new visual indexes are required upon a new object's appearance (referred to as the abrupt onset effect) which is when this shift of attention occurs (Yantis & Jonides, 1996). However, it was unclear if this effect would always occur. This concept was investigated in depth in a study where four letters were arranged on the vertices of a hexagon. Participants were asked to determine which letters (either E or H) was present. On each trial, an arrowhead cue indicated the correct location of the required letter (Yantis & Jonides, 1990). The efficacy of a spatial cue was manipulated by making it appear before, simultaneously or subsequent to the presentation of a test display (Yantis & Jonides, 1990). It was found that endogenous pre-cues promoted highly focused attention and eliminated the abrupt onset effect. Yantis and Jonides (1990) concluded that the abrupt onset effect is attenuated if an individual is engaging in a highly focused attention activity.

Franconeri, Hollingworth, and Simons (2005) examined a more recent view called the transient hypothesis which suggests that types of luminance and motion transients is what captures attention regardless of whether there is a new object or not. For example, if a teacher is reading a story book to a class of children sitting on the floor, if one student stands up; the teacher's attention would be directed to the student. This demonstrates that a new object is not required to capture attention, but instead a luminance and/or motion transient is required. Franconeri et al. (2005) concluded that their experimental results supported the transient hypothesis but did not support the new-object hypothesis. **They found that attention was only captured in new object situations when the object created a unique transient such as luminance or colour change. This suggests that in order to effectively capture the operator's attention, critical information such as a change in airspeed (possibly caused by windshear) should be**

presented in the form of a unique transient. For example, a flashing source of information could capture the operator's attention.

5.5 Concluding Remarks

For multimodal interface designers, there are many factors to consider for the optimal communication of information. The study of multimodal perception, integration and application to interface design not a widely understood field, despite the fact that abundant research exists. However, a number of design guidelines have surface from this review on crossmodal attention:

- Stimuli should be placed in the peripersonal space for maximum effectiveness.

- Auditory and tactile stimuli are best for presenting warning signals.

- Although there is no direct method of prevention for operator confusion by multimodal integration, reference to past experiments and simulation using Bayesian modeling can be useful tools in preventing conflict situations.

- Using multimodal cues may be beneficial in directing the operator's attention to a single location accurately, but the use of multiple senses may slow response time. Thus, the use of multimodal cues depends on the important of accuracy versus response time.

- The use of two sensory modalities is useful when in-parallel processing (attention is spatially divided) is required.

- Low speed stress and load stress result in higher operator performance, thus designers should keep the number of channels and the rate of change of signal presentation low.

- Attention is focused more on channels that update frequently and thus more important parameters should be displayed at a higher frequency.

- Steps should be taken to reduce complacency, which occurs with higher reliable sources. Higher reliable sources should thus be displayed using warning signals instead of monitoring tasks for maximum effectiveness.

6 Intelligent Adaptive Interfaces

Hou, Kobierski, and Brown (2007b) describe *Intelligent Adaptive Interfaces (IAI)* as a system that adjusts the machine's characteristics and/or display to dynamically change with external events in terms of operator states and mission goals in real time. Hou et al. (2007b) have stated that it is an established finding that IAIs can assist in reducing the operator's workload along with contributing to an increase his/her situation awareness. Thus, the IAI domain is an interesting, complex field that researchers have been exploring in attempts to evolve interface designs. Hou et al. (2007b) describe essential qualities of an IAI system include the ability to model human decision making, monitor operator performance, and workload (via behavioural and physiological indications) abilities along with the capacity to predict operator expectations and intentions in relation to the operation's missions, goals, and plans. In the following section various topics that will be addressed include adaptation rules, adaptation guidelines in terms of multimodal displays and existing multimodal adaptive displays.

This section is organized as follows:

- Section 6.1. Discusses adaptation rules and how they are used in an IAI.
- Section 6.2. Describes guidelines for multimodal interfaces.
- Section 6.3. Discusses current implementations of multimodal interfaces with adaptive components.
- Section 6.4. Provides concluding remarks.

6.1 Adaptation Rules

A lack of available guidelines and framework for designing intelligent adaptive interfaces presents many challenges. Thus a brief design framework will be provided in this section. IAIs should adapt to the needs of different users within various contexts. Hou, Gauthier, and Banbury (2007a) provide the following framework for the design of intelligent adaptive systems stating that the combination of the processes below provides a comprehensive framework to develop an IAI (knowledge-based system):

- "*Organization Model*. This model incorporates knowledge relating to the organizational context that the knowledge-based system is intended to operate in (e.g. command and control (C2) structures, Intelligence Surveillance, Target Requisition and Reconnaissance - ISTAR etc.);
- *Task Model.* This model incorporates knowledge relating to the tasks and functions undertaken by all agents, including the operator;

- *Agent Model*. This model incorporates knowledge relating to the participants of the system (i.e., computer and human agents), as well as their roles and responsibilities;
- *User Model*. This model incorporates knowledge of the human operator's abilities, needs and preferences;
- *System Model*. This model incorporates knowledge of the system's abilities, needs, and the means by which it can assist the human operator (e.g. advice, automation, interface adaptation);
- *World Model*. This model incorporates knowledge of the external world, such as physical (e.g. principles of flight controls), psychological (e.g. principles of human behaviour under stress), or cultural (e.g., rules associated with tactics adopted by hostile forces);
- *Dialogue/Communication Model*. This model incorporates knowledge of the manner in which communication takes place between the human operator and the system, and between the system agents themselves;
- *Knowledge Model*. This model incorporates a detailed record of the knowledge required to perform the tasks that the system will be performing; and,
- *Design Model*. This model comprises the hardware and software requirements related to the construction of the intelligent adaptive system. This model also specifies the means by which operator state is monitored."

For an interface to be able to adapt, it must be capable of collecting data on personal features via implicit and explicit behaviour. Examples of personal data include personal preferences, experience, sequential demands, task demands, operator state, physical conditions (e.g. ambient noise level), user aptitudes (e.g. spatial reasoning ability or visual acuity), user demographics, workload etc. (Hameed & Sarter, 2009; Meyer, Yakemovic, & Harris, 1993). Additional aspects to consider is the type of adaptation the interface system will assume. A few types of adaptations include "*task allocation or partitioning* which is when the interface completes the entire task or a portion of it; *interface transformation* in which the system facilitates the task in attempts to reduce the difficulty by adapting the communication style, content and form of displayed information; *functionality* in which the interface changes the available functions depending on user differences; and *user* in which the system assists the user by assuming the role of a tutor" (Meyer et al., 1993). The type of automation and task allocation is a particular design issue that adaptive interface designers encounter. Task allocation between the system and the operator can significantly influence the effectiveness of the interface and the user's experience. Sheridan (2000) argues that the optimal level of automation varies at different stages of a task. For example, the operator should be an active participant rather than a passive monitor. The assignment of humans taking on the role of monitoring performance has been indicated as a design weakness in current automation interfaces. Using humans as a resource to monitor performance is considered a weakness because humans often encounter detrimental factors such as boredom and vigilance while engaging in activities that require monitoring performance (Sherry & Ritter, 2002).

The following points are additional rules that should be taken into consideration:

- Humans are held responsible for the task's overall performance therefore they must be given control authority (Sherry & Ritter, 2002).

- The information presented (i.e. in displays) should be hierarchically organized. By employing a hierarchy system for goals and tasks, interfaces can support the user's activities effectively.

- "The operator interface functional requirements and associated IAI to be incorporated should be fully described before the rapid-prototyping software effort starts. Subsequent interface concepts may require significant changes to the core software structure and will be resisted by the software engineers once they have invested time in a preliminary architecture" (Hou et al., 2007b).

- Hou et al. (2007b) stated "an IAI could be designed to make recommendations then take appropriate actions according to a "yes" or "no" or "implement without asking" operator response. At some point however, the operator could request that the IAI not make recommendations in a certain area but, rather complete the IAI suggested action without user input (full automation)."

- Hou et al. (2007b) stated that IAIs "should include a feature that allows the operator to return to the previous system's state prior to an IAI automatic configuration/task. It is also important that the IAI inform the user of all functionalities and decisions that the system assumes."

6.2 Multimodal Adaptive Display Design Guidelines

Society is striving for the most efficient, easy and flexible form of interaction with interfaces for information retrieval (Croft, 1995). Thus, implementing adaptive multimodal interface designs seem to be a possible method of improving interfaces for efficient information retrieval. Information input to systems can also be expanded beyond normal keyboard/mouse interactions to other multimodal input methods such as voice and gesture, but the focus of this section, as it was for the whole report, is on multimodal output. Attempts for literature review of adaptation guidelines to multimodal displays have been conducted; however, concrete guidelines for this area have yet to be established. Nonetheless, basic guidelines have been provided. In order to convey information adaptively in a multimodal way, the following six considerations must be taken into account:

- Choice of the information that is to be conveyed ("content selection").

- Selection of modalities through which the information will be conveyed ("modality allocation").

- Selection of the format in which the modalities will be used to present that information ("modality realization").

- Determinations of mechanism(s) that are used combine the modalities ("modality combination").

- Evaluating the effect of environmental and cognitive factors on user's perceptual integration ("situated multimodality").

- Analysis of performance of the human user in the interface ("task analysis") (Tripathi, 2008).

As mentioned in the previous section above, the interface's ability to collect personal data (individual differences) allows the display to present information adaptively and in accordance with the situation and the user's needs and preferences. For example if the interface detects that an operator is experiencing an overload in the visual modality, the interface could adapt and present information through another modality such as tactile or audition. Hou et al. (2007b) pointed out that it is absolutely vital for the operator's states and intentions to be clear to the interface; thus, it would be helpful for the interface to indicate its perception of the operator's states, intentions and mission goals. Additional basic multimodal guidelines are as follows:

- "Maximize advantages of each modality to reduce user's memory load in certain task and situations

- Integrate compatible modalities in context with user preferences and system functionality for example, allow gestures to augment or replace speech input in noisy environments

- Avoid presenting information in different modalities unnecessarily in cases where the user must attend both sources to comprehend the material being presented. This can cause an increase cognitive load at the cost of learning material" (Reeves et al., 2004).

- Selection options for preferred presentations via different modalities should be available

- Users should be able to adjust in terms of scalability individual modalities. For example, features within individual modalities such as display contrast should be able to adjust in accordance to the environment and the user's preferences

- Schneider-Hufschmidt, Groh, Perrin, Hine, & Furner (2003) said that information content should be designed appropriately to provide constant multimodal presentation and be stored in "delivery-independent form" so that translations of information in different modalities are consistent. This statement contradicts the first guideline provided within this list. The first statement appears to be a more intuitive guideline since attempting to present information "delivery-independent form," may result in downplaying modality specialization. It is important that the development of multimodal adaptive interfaces select modality usage optimally.

- Modality selection in the design stage should be determined by two factors: appropriateness and availability in relation to various factors such as urgency, purpose, information importance, and processing code along with each modality being assigned a rank order value of 0-1 depicting its desirability level; 0 being the least desirable and 1 being the most desirable (Hameed & Sarter, 2009). It is not specified why the authors suggested a ranking system from 0-1 for desirability level; however, another ranking system (e.g. 1-5) could also be employed as long as each ranking level is clearly defined and consists of specific criteria. This allows the interface to take on the responsibility of automation in accordance with the user's preferences and needs by collaborating authority amongst the user and the interface.

For the most part, guidelines of multimodal adaptive interfaces appear to be consistent across the literature. However, there appear to be contradicting guidelines. One guideline states that user preferences of modalities should exist while another states that modality availability should be used in relation to its availability and appropriateness. The best solution seems to be that interfaces should only provide different modality display selections if the information can be presented through different modalities optimally.

6.3 Existing Adaptive Multimodal Displays

There are very few existing adaptive multimodal displays. Our literature revealed two systems that are described here, the Gaze-X system (Maat & Pantic, 2006), and an online learning system by Pentland and Roy (1998). Although both systems use multimodal information and adaptive interfaces, both systems suffer significant limitations that prevent them from providing much insight into adaptive multimodal interface design.

The Gaze-X is a multimodal display interface that models the user's emotions and actions and in return adapts the interface to support the user's activity. Gaze-X uses multimodal input as a framework for adaptation. It can process the user's facial expression, eye gaze direction, speech, keystrokes and mouse movements and actions such as pointing to an object (Maat & Pantic, 2006). This system operates within the context referred to as the "W5 – who, where, what, when, why, how". Questions that the interface derives from the user's emotions and actions are "who is the user? Where is the user? What is the current task of the user? How the information is passed on? Which interactive actions/signals were used? When is the timing of displayed interactive signals? Why the user chose to display the observed cues?" (Maat & Pantic, 2006) This multimodal display follows all the guidelines provided above. For example, the user can disable automation functionalities and is able to change the modality of the information presentation. One problematic area with this adaptive display is the system's method to detect the operator's mood state. This system uses a web-cam, face reading system that can detect prototypic facial expressions based on six different emotions which are surprise, fear, sadness, disgust, anger, and happiness (Maat & Pantic, 2006). An issue with this method is that the operator is relying on solely external visual information to interpret the operator's mood state. Implications can arise in situations where the system does not correctly interpret the user's mood. For example, some individuals laugh when they are nervous for various reasons such as fear and anxiety. In this situation, the Gaze-X would interpret the user as happy and function in accordance to that (i.e. provide less assistance). Thus perhaps a more effective way to interpret the user's state is by combining internal with external readings. For example EEG and fMRI readings could be used in parallel with the face reading system implemented by Gaze-X to determine the operator's state.

Another adaptive multimodal interface developed to act as an on-line learning tool that allows the user to communicate with the system through speech and deictic (e.g. pointing) gestures (Pentland & Roy, 1998). This system employs a vision based hand tracking system and a speech recognizer along with an animated character, Toco the Toucan who is referred to as "Toco." Toco learns to associate words with objects when users associate one with the other. The reason for this feature is due to users referring to the same object in various ways and/or meanings (Pentland & Roy, 1998). For example, if the user said “Toco” and pointed to an object and said “ball,” Toco would associate the word ball with that specific object. An issue with this multimodal adaptive interface is that it seems that the user must spend time teaching the interface instead of vice versa. Additional issues include a lack of conforming to the guidelines above. For example, the information communicated with the system is not transferrable across modalities (i.e. if the user is occupied verbally or physically and is not able to say the object’s name or point at the object – the learning goal for this interface will be disabled). Although this interface’s purpose is to train the system to recognize new objects, it appears that this is just a preliminary design and the authors have future goals such as requiring Toco to perform actions with or to the objects. This is one example of a multimodal input interface, through the use of gestures and voice, which attempts to be adaptive by learning new associations.

6.4 Concluding Remarks

Overall, many design guidelines and concepts have been provided in terms of multimodal adaptation that should be carefully addressed and considered prior to the design of interfaces. In addition, existing adaptive multimodal displays have been mentioned and can be used as exemplars to the interface this project is focusing on but there are very few such implementations and they have all been quite limited in scope.

Adaptive multimodal interface design must first conform to the guiding principles of good multimodal interface design. Further to the design of these interfaces, research must generate more principles for adaptive multimodal design. Hameed and Sarter’s work (2009) suggest that the adaptive presentation of urgent and important information in the multimodal domain would be the most productive first direction to explore. Ideally the information should complement the other modality information to avoid having the user process two modalities simultaneously. Choosing the adaptation triggers is critical to this research but a few potential directions could be:

- Workload as measured through physiological response.
- Visual loading/attention loading/or cognitive tunnelling as measured through lack of fixation on critical information, possibly using eye-tracking.

7 Developing a Program of Research

In this section we review information that is directly relevant to the design of our experiment which makes use of the GCS interface being developed by DRDC Ottawa. Our focus, as it was in the rest of this report, is on supporting the development of an enhanced interface for supporting the task of monitoring UAV landings. As such, we do not provide recommendations, outside of those already addressed by EID, for the baseline GCS interface. This interface is based on current UAV GCS interfaces that make very little use of non-visual information presentation.

This section is organized as follows:

- Section 7.1. Reviews the use-cases for the autoland abort scenarios that can be modelled by the UAV simulator, and provide a discussion on the cognitive loads imposed by each scenario.

- Section 7.2. Examines literature that is relevant to the UAV autoland monitoring scenario and identifies methodologies that can be adapted for use in future studies.

- Section 7.3. Proposes new lines of experimentation based on the literature covered previously in this report.

7.1 Cognitive Task Loading of UAV Autoland Scenarios

This section is intended to outline perceived cognitive workload of pilots of manned aircraft as a parallel to the actions of operators of unmanned aerial vehicles (UAVs), specifically for medium-altitude, long-endurance UAVs. This re-assessment of the cognitive workload was done from the perspective of an experienced airline pilot.

This section outlines approximated cognitive workloads based on the following criteria:

- ***Timing*** – if timing is tight or timing is relaxed/not a concern
- ***Accuracy*** – actions need to be executed accurately/there is a reasonable buffer for error
- ***Information*** – there would be lots of information coming in to the operator/there would be fairly little information
- ***Computation*** – the pilot would have a lot of things to evaluate and assess/there would be very little mental workload

- ***Memory*** – there would be a lot of memory required/there would be very little memory required

- ***Training*** – this scenario would be very hard for novices/this scenario should be easy with a low level of training

- ***Depth of the action chain*** – this scenario requires many steps to complete/this scenario is essentially a single step.

Each use case is then examined using these parameters. The intent being that, by using these cases with respect to manned aircraft operators (pilots), a clearer and more focused framework of cognitive task load can be approximated to the UAV operator. Every effort was made to give the more cursory experimenter a richer understanding of the cognitive workload of aircraft pilots as they deal with the following scenarios

7.1.1 Case I – Low Fuel Abort

The low fuel abort manifests itself with a visual indication (amber or red) indicating to the operator the amount of fuel remaining in the aircraft (this indication can be either in time (i.e. – XX Minutes of flight remaining) or, as is common with manned aircraft, amount of fuel measured in either pounds or litres.

- ***Timing*** – This condition (either red or amber) requires the operator to perform a number of functions but timing is not necessarily critical. Reaction time to this scenario can be measured in minutes and the amber warning could be measured in tens of minutes.

- ***Accuracy*** – Accuracy calculation can be important in this scenario if the result is to continue to an appropriate airfield. The accuracy of continuing to destination is explained further in computation

- ***Information*** – The amount of information the operator receives is general low. In manned aircraft the only indication the pilot would initially receive is an amber warning indicating the amount of fuel has reached a pre determined caution stage. Normal pilot reaction would be to confirm the amount of fuel is on the aircraft in order to determine that the warning system has not generated a false positive. If the indication is true (there is low fuel) generally no other indication will be present until the fuel reaches a predetermined second stage, in which the amber warning will become red. This red indication, describing the dire fuel condition of the aircraft, still equates to minutes of flying time (around ten minutes).

- ***Computation*** – From a computational standpoint, the pilot needs to be able to determine if the fuel remaining is sufficient to continue on to the final destination or if alternative requirements are needed. Simple mental math calculations are the most that are required. For example

    - *If the aircraft has 100 lbs of fuel remaining and the aircraft fuel flow is 200 lbs of fuel per hour the aircraft has approximately (100 lbs / 200 lbs / hr = 0.5 hr) 30 minutes of fuel remaining. Given the fuel state, if the aircrafts groundspeed (the* *speed it makes over the ground based of the prevailing winds at the present time) is 250 mph and the distance from destination is 500 miles, the aircraft will not make the destination (250mph*.5 = 125 miles).*

- Given this particular scenario, alternative landing fields could be determined in coordination with other participants however the integrity of the UAV is not in immediate danger

- ***Memory*** – Memory requirements are as described in the computation section of the scenario

- ***Training*** – Training is based on recognition of the fuel state and being able to determine a next course of action. There is little in the way of systems or technology training required for this case study

- ***Depth of the action chain*** – The action chain in this case is quite long. Based on the scenario outlined in the computation scenario, actions start with the determination of "Is the UAV going to make it home?" If yes, the decision making loop is closed. If not, the required actions include re-evaluation of the decision to continue on to destination. For example "is my determination that the UAV will have enough fuel to reach the destination" hold true? If it does, the decision stands and the flight can continue. If after deciding to proceed to destination, fuel consumption calculations show that there will be not enough fuel to reach, a new course of action (COA) is required. If a forced landing is then required, consultation to determine best destination may require input from various sources (i.e. – can we land it near friendly troops in order to retrieve the UAV).

7.1.2 Case II – Power Bus Related Abort

As in the case of the low fuel scenario, a Power Bus related abort requires a determination of the severity of the situation. The situation (amber or red indication) will determine the COA to take.

- ***Timing*** – This condition (either red or amber) requires the operator to perform a number of functions but timing is not necessarily critical. Reaction time to this scenario can be measured in minutes and the amber warning could be measured in tens of minutes

- ***Accuracy*** – Accuracy calculation can be important in this scenario if the result is to continue to an appropriate airfield. The accuracy of continuing to destination is explained further in computation

- ***Information*** – Information can be more in depth than the fuel scenario. In the case of bus loss, certain systems will be lost. In manned aircraft (particularly transport aircraft) the loss of a systems bus is generally a non issue as a supplemental power source, the auxiliary power unit (APU) can provide supplemental power in the event of a bus loss in flight. The UAV does not have such a system, therefore a bus loss will be indicated by various systems losses and an amber or red warning. The red warning, in case of electrical discharge will indicate a certain amount of time remains (about 10 minutes) before all electrical power is lost. In the case of manned aircraft, a red warning requires an immediate forced landing at a suitable landing site

- ***Computation*** – From a computational standpoint, the pilot needs to be able to determine a suitable field to land in case of a red warning. This requires a determination of the local topography and estimated landing distances. For example, as is the case with low fuel

    - *If the local topography is mountainous, an immediate landing may not be available. A quick determination of the distance to appropriate landing areas may be required and a determination of distance versus time remaining will need to be computed.*

  - Given this particular scenario, alternative landing fields could be determined in coordination with other participants however the integrity of the UAV is not in immediate danger

- ***Memory*** – Memory requirements are as described in the computation section of the scenario

- ***Training*** – Training is based on recognition of the aircraft state and being able to determine a next course of action. Training for an electrical problem requires knowledge of the aircraft standard operating procedures (SOPs). Within the SOPs are steps to reduce electrical loading in order to preserve output. With the checklist complete, potential landing scenarios can be formulated.

- ***Depth of the action chain*** – As in the case of a fuel scenario, the action chain for electrical problems can be quite long. Based on the scenario outlined in the computation scenario, actions start with the determination of "Is the UAV going to make it home?" If yes, the decision making loop is closed. If not, the required actions include re-evaluation of the decision to continue on to destination. For example "is my determination that the UAV will have enough time to reach the destination" hold true? If it does, the decision stands and the flight can continue. If after deciding to proceed to destination, time calculations show that there will be not enough time to reach the destination, a new course of action (COA) is required. If a forced landing is then required, consultation to determine best destination may require input from various sources (i.e. – can we land it near friendly troops in order to retrieve the UAV)

7.1.3 Case III – Windshear Abort

Windshear is a condition caused by frontal air activity. It is defined as a change of windspeed and direction over a relatively small area (1-2 miles). These changes can include airspeed fluctuations of +/- 30 knots from the prevailing winds.

Strong outflow from thunderstorms causes rapid changes in the three-dimensional wind velocity just above ground level. Initially, this outflow causes a headwind that increases airspeed, which normally causes a pilot to reduce engine power if they are unaware of the wind shear. As the aircraft passes into the region of the downdraft, the localized headwind diminishes, reducing the aircraft's airspeed and increasing its sink rate. Then, when the aircraft passes through the other side of the downdraft, the headwind becomes a tailwind, reducing airspeed further, leaving the aircraft in a low-power, low-speed descent. This can lead to an accident if the aircraft is too low to effect a recovery before ground contact.

- ***Timing*** – This condition (different to the previous ones), requires immediate action from the operator. The reaction time here, as opposed to the previous scenarios, is indicated by seconds as the immediate destruction of the aircraft or UAV is more likely without the direct intervention of the operator to remove the aircraft from the windshear condition.

- ***Accuracy*** – Once the windshear condition is recognized, the operator is required to perform the Windshear escape manoeuvre in order to extricate the UAV from the prevailing condition. The danger here is that if the operation is not performed, the aircraft may impact the ground.

- ***Information*** **–** Depending on the interface, most manned aircraft has a windshear detector as a standard form of equipment. There are two types of systems. A normal windshear system will provide aural warnings to the pilot. This aural warning is in the form of "windshear, windshear, windshear" and any properly trained pilot would react to this warning be immediately performing the windshear escape manoeuvre. (See training for definition). Another type of windshear indicator is a predictive windshear warning system. This is characterized by an aural warning that pre-empts the windshear condition by announcing "Caution windshear ahead, Caution windshear ahead". The intent being the pilots can avoid the windshear altogether.

  - If the aircraft does not have windshear detection, it is the operator's task to be able to recognize the preconditions of windshear and avoid it (training element). If it is encountered, the present aircraft instruments can indicated the active windshear but it the pilot's responsibility to recognize and correct the flight path of the aircraft. Information required will be the descent rate of the aircraft (anything in excess of 2,500 feet per minute is an indication of the presence of windshear). As well, excessive pitch and airspeed changes (both increasing and decreasing airspeed) without input from the operator should be an indicator that a windshear condition may be occurring.

- ***Computation*** – There is little in the way of computation in this scenario as it is characterized by a "if-then" logic (i.e. If the windshear warning is activated, then do the escape manoeuvre)

- ***Memory*** – This, in conjunction with training, makes up the core response to the Windshear scenario. If the aircraft has a windshear detector, the operator must remember the actions to be taken in order to escape the condition. In the case of manned aircraft in a landing scenario, the reaction would be

  - ***Maximum power***
  - ***Pitch the aircraft up to a maximum of 20%***
  - ***Fly just above the stall warning***
  - ***Leave the aircraft in its present landing configuration (flaps and landing gear)***
  - ***Once the windshear is no longer a threat***

    - ***Configure the aircraft is if a normal takeoff has just been accomplished***
    - ***Complete the after takeoff***

- ***Training*** – There are training components to the windshear scenario. The first is knowledge of meteorology; the second is an aircraft SOP understanding.

    - Training in meteorology requires the operator to be able to recognize the weather conditions that could provide possibility to the presence of windshear. For example a frontal passage can provide the pre cursors required for windshear (i.e. a cold front passage that produces thunderstorms). If the temperature difference between the warm and cold air mass exceeds 9C then this is another indicator that windshear is likely

    - SOP knowledge requires that upon recognition of the windshear, the pilot executes a Windshear escape manoeuvre (as described under memory).

- ***Depth of the action chain*** – Upon recognition of, and reaction to, the windshear condition, the action chain is relatively short as the windshear phenomena is very localized and will not last for more than a few seconds. It is described as an "if-then" reaction. If in windshear then perform the escape manoeuvre. Once clear of the windshear, the decision is made to either wait for the weather to pass (usually not more than a few minutes). Or proceed to an alternative landing site.

7.1.4 Case IV – Excessive Vertical Velocity Abort

The cognitive load for windshear includes the specific loading for excessive vertical velocity and is not considered as a specific scenario.

7.1.5 Case V – Excessive Pitch Abort

As was with case IV, Windshear abort includes the cognitive workload of the Excessive pitch abort and is again not considered as a specific scenario.

7.1.6 Case VI – Wing Icing Related Abort

In- flight icing is a serious hazard. It destroys the smooth flow of air, increasing drag, degrading control authority and decreasing the ability of an airfoil to lift. The actual weight of the ice on the aircraft is secondary to the airflow disruption it causes. As power is added to compensate for the additional drag and the nose is lifted to maintain altitude, the angle of attack increases, allowing the underside of the wings and fuselage to accumulate additional ice. Ice accumulates on every exposed frontal surface of the aircraft – not just on the wings, propeller, and windshield, but also on the antennas, vents, intakes, and cowlings. It builds in flight where no heat or boots can reach it. It can cause antennas to vibrate so severely that they break. In moderate to severe conditions, a light aircraft can become so iced up that continued flight is impossible. The aircraft may stall at much higher speeds and lower angles of attack than normal. It can roll or pitch uncontrollably, and recovery may be impossible.

In the landing regime, a significant amount of accumulated ice on airplane increases weight. This increases stall speed and as such, the pilot is expected to increase the approach speed. An increase in speed will provide a better margin above the stall speed. If the pilot detects significant amount of accumulated ice on landing approach, then the pilot should attempt to land. The assumption that the de-icing equipment is capable of clearing all the ice to facilitate an abort should not be made.

- ***Timing*** – Icing conditions require various timings based on the severity of the icing and the phase of flight being performed. In the landing phase, a COA needs to be determined rather quickly because of the phase of flight characteristics. The aircraft is in a slower speed regime based on the decision to land and the aircraft is low to the ground. This low altitude restricts the operator from having the time to resolve any performance issues if the icing becomes too great.

- ***Accuracy*** – Once the icing condition is recognized, the operator is required to determine the COA. Does the aircraft continue to land or is it prudent to perform a missed approach and attempt to land at an alternative landing destination. The danger here is that if the operator performs a missed approach, the aircraft may prolong its stay in icing conditions to such a point that physical flight becomes impossible.

- ***Information*** – Depending on the interface, most manned aircraft have some type of icing detector. This would be in the form of an aircraft sensor that shows up in the flight deck or pilot console. As icing itself is not a grave danger (aircraft fly through icing conditions every day) the detector may be a caution or amber light. It is important to note that the degree of icing is not indicated to the activation of the caution light. Light icing or severe icing gives the same message. It is the pilot's responsibility to determine the degree of icing occurring on the airframe. With a manned flightdeck, this is relatively easy as aircraft have other devices, such as dedicated icing displays, which can be physically observed to determine the rate of ice accumulation. A remote pilot will need to determine the next COA based on limited information. The presence of icing conditions or whatever information can be gathered through the on board camera.

- ***Computation*** – There is little in the way of computation in this scenario as it is characterized by a "if-then" logic (i.e. If the icing warning is activated, then activate de-icing equipment)

- ***Memory*** – If the decision is to land, then no other reactions are required other than a speed increment. Icing conditions require the aircraft increase speed in order to counteract the increased stall speed of the contaminated wing. In manned aircraft this is manifested by some standard speed factor (i.e. If in icing conditions, increase the landing speed by 10 knots)

- ***Training*** – As with the windshear scenario, there are training components to the icing scenario. The first is knowledge of meteorology; the second is an aircraft SOP understanding.

- Training in meteorology requires the operator to be able to recognize the weather conditions that could provide possibility to the presence of icing conditions. For example a warm front passage can provide the pre cursors required for airborne icing.

- SOP knowledge requires that upon recognition of the icing, the pilot executes a missed approach and conduct an after-takeoff checklist

- ***Depth of the action chain*** – Upon recognition of, and reaction to, the icing condition, the action chain is relatively short as the icing phenomena is very localized. If the decision is to land, there are no additional factors other than an incremental speed increase. If the choice is to attempt another landing, a go-around procedure (identical to a take-off procedure) is required. At the completion of the take-off, the next COA is to determine if the aircraft should wait until the condition has passed (unlike windshear, this could be a much longer wait and could cause fuel concerns), attempt another approach (with the possibility of the same condition happening again) or determining an alternative landing location.

- The preceding cases outline the more cognitive aspects of performing each of these manoeuvres. In some cases, there was little in the way of active cognitive performance. Others provided for more interpretation and decision making as opposed to the “if-then” statements.

7.1.7 Case VII – Engine Health Related Abort

Engine health related aborts could be many different problems requiring really only one solution – the immediate landing of the aircraft. However, it is important to note that essentially the aircraft is still flyable. Air Transat proved that an Airbus A330 (approximately 350,000 lbs) could safely glide over 100 miles without *any* engine power. We can extend this to the use case of the engine health issue.

The Heron is a single engine UAV. If the aircraft follows larger manned aircraft design, its engine could also provide auxiliary power to various other systems by an accessory gear box. If the engine stops running, so do many other seemingly unrelated systems. Therefore, in order to maintain other systems, the aircraft’s engine needs to be running.

- ***Timing*** – This is not a critical item as timing can be measured in minutes. Engine health issues require checklist reviews and an assessment of the requirements to recover the vehicle. Essentially any caution or warning alarm is addressed by a checklist to completion. Once completed, an assessment as to landing needs to be made. Single engine manned aircraft leave the aircraft engine running and are not concerned with the health of it. In comparison, multiple engine aircraft would shut the faulty engine down. In order to be certified to carry passengers, all multiple engine aircraft must be able to safely fly in all flight regimes with the use of only a single engine.

- ***Accuracy*** – The only accuracy required is to be able to complete the appropriate checklist and properly assess the craft situation in order to recover it.

- ***Information*** – The first information required is the type of engine health issue. In large aircraft, these could be oil temperature and pressure indications, thrust reverser deployment (jet engines) or propeller governor failure (propeller driven engines). It should be noted that some engine health issues may not require the engine to be shut down but rather to be run at a slightly retarded thrust setting. For example, high oil pressure would require a retarded thrust setting in order to reduce the pressure. The engine could be operated at a lower setting continuously until the craft is recovered. Once the state of the aircraft is determined and established, the subsequent information would be the decision to continue the mission or recovery.

- ***Computation*** – There is very little computational issues in this scenario. If the UAV was a multiple engine craft, this would not be the case as an engine shutdown would require speed and fuel flow re-computations.

- ***Memory*** – This requires very low memory as engine health issues (whatever the issue) is normally dealt with using a checklist.

- ***Training*** – The training scenario for this use case is high. Use of checklists requires some practice in order to use it in scenario.

- ***Depth of the action chain*** – The action chain is long in this scenario. There are several stages. The assessment and checklist required for the health issue and the decision to continue or recover the vehicle. In large aircraft, checklist and final destination could take several minutes. Performance issues in recovery also need to be considered (can the craft fly over certain obstacles)

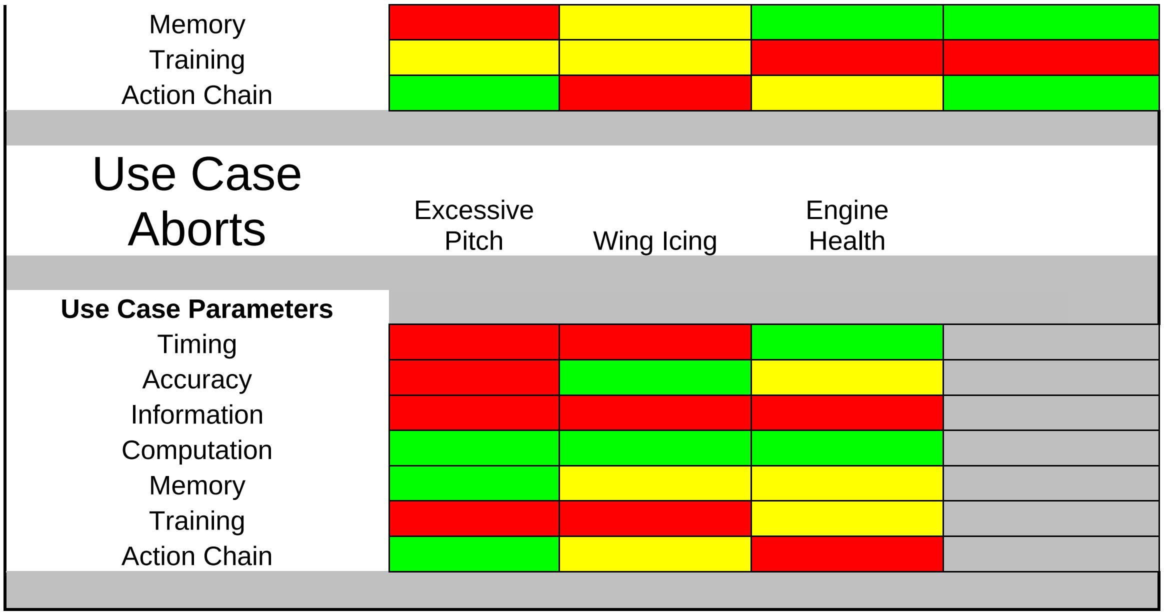
    -

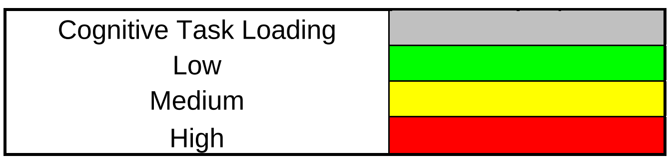
7.1.8 Summary

As a conclusion, all use cases are summarized in the following chart. Use cases and cognitive workload are compared and an assessment of their relative cognitive loading are presented (Low / Medium / High).

Table 8: Cognitive Loading for Use Case Abort Scenarios

Use Case Aborts	Low Fuel	Power Bus	Windshear	Excessive Vert. Vel.
Use Case Parameters				
Timing				
Accuracy				
Information				
Computation				





7.2 Experimental Methodologies

In this section we discuss examples of experiments that have high relevance to our current research in respect to methodology or application. The following table summarizes the major elements of each of these relevant papers. A more thorough discussion of each paper can be found in the appendix summaries. (Calhoun et al., 2003; Calhoun et al., 2004)

Table 9: Relevant Experimental Methodologies

Paper	Application	Modalities	Experimental Platform	Primary Task(s)	Secondary Task(s)	# Ss
Burns (2000)	Power Plant (Process Control)	Visual	Simulation	Fault Diagnosis	N/A	18
Kramer et al. (2000)	Commercial Aircraft (Autoland)	Visual	Simulation	Autoland Monitoring	N/A	8*
Calhoun et al. (2003)	UAV (Flight Monitoring)	Visual, Auditory, Tactile	Simulation	Tracking (Flight)	Check List Tasks	10
Calhoun et al. (2004)	UAV (Flight Monitoring)	Visual, Auditory,	Simulation	Tracking (Flight)	Check List Tasks,	12

(experiment 2)		Tactile			Auditory radio task (CRM)	
Aretz et al. (2006)	UAV (Landing/Training)	Visual, Tactile	Simulation	Tracking (Flight)	N/A	30
Brill et al. (2008)	N/A	Visual, Auditory, Tactile	Lab Experiment	MATB	M-SWAP	31
Donmez et al. (2008)	UAV (Flight Monitoring)	Visual, Tactile	Simulation	UAV supervisory control	N/A	13
Oskarsson et al. (2008)	Combat Vehicles (Threat orientation)	Visual, Auditory, Tactile	Simulation	Threat Orientation	Auditory radio task	12
Tadema and Theunissen (2008)	UAV (Autoland)	Visual	Simulation	Autoland Monitoring	N/A	52
Maza et al. (2009)	UAV (General)	Visual, Auditory, Tactile	Lab Experiment	Spatial Discrimination and Response Task	N/A	9

* denotes trained or professional participants (e.g. pilots)

7.2.1 Primary Tasks

A variety of primary tasks existed in the literature reviewed and the type of task used was largely dependent on the domain of application. The primary task was defined as the task that participants were asked to focus on, or the one that had the highest priority if there were multiple tasks presented. Two major groups of application tasks were supported by the research: those which involved manual control (e.g. Aretz et al., 2006), and those which involved monitoring and human-supervisory control (e.g. Burns, 2000). A review of the above literature showed that many of the studies involved some aspect of both types of tasks. However, manual control tracking tasks were used more often as the primary experimental task, while the monitor tasks were regulated to being secondary tasks.

7.2.1.1 Tracking Tasks

The most common tracking task used was adhering to a preset path during flight. Calhoun et al. (2003; 2004) used a UAV monitoring task where the participants were asked to maintain an altitude and airspeed while flying along a path in an UAV simulator. Participants were presented with stimuli which closely matched those found in current GCSs that require manual control (display with map and other mission relevant data, display with simulated video imagery from a nose camera with additional overlays, a third display with subsystem and communications information, a control stick, and a throttle control), and they relied solely on this visual information to accomplish the tracking task. In order to accomplish the tracking task, the participants had to keep track of the location of their UAV on their map display and the UAV’s current altitude and airspeed while using the control stick and throttle control to keep the UAV within the required boundaries.

In another experiment, Aretz et al. (2006) had participants manually land an UAV. Participants had controls and visual displays that were similar to those found in the Calhoun studies. However, participants were also presented vibrotactile feedback for altitude deviations through a tactor vest. Tactile feedback was provided via a tactile vest with four rows of tactors. Each of the rows represented different levels of deviation from the optimal altitude during the approach. The top most row would vibrate intensely (200ms on, 100ms off) if the UAV was 20 feet above the optimal glideslope. The second highest row would vibrate softly (100ms on, 600ms off) when the UAV was 10 feet above the optimal glideslope. A similar coding strategy was used for the bottom two rows for when the UAV was below the optimal glideslope. Kramer et al. (2000) also used a landing scenario where participants were required to hand fly a simulated aircraft through the last stages of an approach when automation failed. This was done to measure the participant’s situation awareness (SA) through the use of screen blanking. Participants who were previously required to do an aircraft autoland monitoring task would sometimes be confronted with a scenario that involved blanking their data displays, simulating automation failure while also removing sources of information. The participants were then required to fly the rest of the approach using only a single “back-up” instrument. Their ability to do so would be related to how situationally aware they were before the blanking occurred.

The most common dependent measure used with tracking tasks is root mean square (RMS) error. While tracking was used quite often as a primary task, the independent variable manipulations were rarely done to affect changes in tracking behaviour (Aretz et al. is one notable exception). Instead, the tracking task was used as a loading task, since many of the experiments were interested in supporting tasks during high workload conditions. Also, all the tracking tasks which have been discussed are visual tracking tasks. While it was never explicitly stated in any of the papers, good performance during the visual tracking task would require attentional resources to be directed to the visual modality. It is also safe to assume that the participant’s attention is also spatially focused on the relevant monitors.

7.2.1.2 Monitoring Tasks

Many types of monitoring tasks were used in the experiments listed above, but a common element in these tasks was that they all involved observing a large number of information channels for specific events or conditions. In an experiment by Burns (2000), participants were asked to monitor an interface displaying information about the water cycle in a coal-fired power plant. Faults and problems were introduced into the power plant simulation at random times in the simulation, and participants had to detect these faults as quickly as possible. In addition to detection, Burns also had participants diagnose the cause of the fault or problem. The additional diagnosis step made this a cognitively difficult monitoring task because participants were required to problem solve and integrate different pieces of information. Donmez et al. (2008) had participants monitor the progress of four UAVs simultaneously as they completed automated missions. Participants were told to monitor and correct for course deviations (only when the UAV reached a certain threshold of deviation), and respond to late arrivals (when an UAV is unable to reach a waypoint at the scheduled time) based on a set procedure. This was an example of a perceptually difficult monitoring task because it required that the participant focus on many different spatial locations. Tadema and Theunissen (2008) conducted a study on how synthetic vision overlays can improve an operator's ability to supervise an UAV autoland scenario. Participants were required to assess the integrity of the guidance information used by the autoland system during the approach while using an interface with or without synthetic vision overlays. Participants could either allow the UAV to land or they could instruct the UAV to go-around during each approach. In another study which examined autoland monitoring, Kramer et al. (2000) measured the SA of participants using different types of visual interfaces. Similar to Burns (2000), they used an "Anomalous Cue/Detection Time" technique, where they introduced a problem into the simulation and measured the time until detection and diagnosis. Faster response times would imply higher levels of SA.

Monitoring tasks tend to use accuracy and response time as dependent measures. Both are measured to ensure that speed-accuracy trade-offs are not occurring. Also, monitoring tasks often require the use of pre-planned scenarios because the goal in a monitoring task is to detect when a set of conditions are met. Since pre-planned scenarios are often discrete and have a single correct answer, signal detection analysis (such as in Burns, 2000) can also be used on monitoring tasks. While tracking tasks were normally used as a loading task, the monitoring tasks used in the experiments above were designed to be affected by manipulations of the independent variable. However, Brill et al. (2008) used the Multi-Attribute Task Battery (MATB), where participants had to monitor four horizontally arranged bars with a moving pointer. The participants were required to monitor for "malfunctions" based on the values of the bars, and respond by hitting a button. Brill et al. were interested in loading the visual modality so that a secondary task (M-SWAP) could be used to measure reserve cognitive capacity in different modalities.

For monitoring tasks that involve classification or diagnosis, it may also be important to gauge the degree of correctness of the participant's answer. Burns (2000) used a 4-point ordinal scale to show the accuracy of the participant's fault diagnosis. Missed diagnoses were rated 0, diagnoses that only referred to symptoms (but not the higher level cause) were rated 1, correct but vague diagnoses were rated 2, and a completely correct diagnoses was rated 3. By assessing the correctness of a diagnosis, the experimenter is able to discover why a participant made the classification that was chosen.

7.2.2 Secondary Tasks

Three different types of secondary tasks were used in the papers reviewed in this section: the Multisensory Assessment Protocol (M-SWAP), auditory "radio" secondary tasks, and check lists.

Brill et al (2008) examined the plausibility of independent pools of resources for different modalities through the use of the M-SWAP secondary loading task. M-SWAP is a secondary task measure which makes use of perceptual signals in different modalities to gauge reserve cognitive capacity. Each perceptual signal was composed of three possible channels of information. For example, the visual signal consisted of three white boxes. During each stimuli presentation, one of the channels would be activated. Similar signals were constructed for the auditory modality (three tones at different frequencies) and the tactile modality (no specific description of the tactile modality was included in the paper). Participants were asked to monitor a specific channel and note how many times stimuli was presented through that particular channel. Each time the participant counted four stimuli presentations in the observed channel they pressed one of three response buttons. The difficulty of the task could also be increased by asking the participant to monitor multiple channels at once. The dependent measure was the number of counting errors made.

Two studies used some variant of a radio-based auditory secondary task. Calhoun et al. (2004) used a modified version of the Coordinate Response Measure (Bolia, Nelson, Ericson, & Simpson, 2000 as cited by Calhoun et al., 2004). Radio calls, composed of a call sign, a colour, and a number (e.g. ready *Eagle*, go to *blue 8*) were played, and participants were required to respond to radio calls that were directed to their callsign and conduct a data entry task based on the colour and number in the radio call. Calhoun et al. also manipulated the difficulty of the auditory task by having only relevant callsigns for the low auditory load condition, and by having 8 different callsigns for the high auditory load condition. A manipulation check for auditory load (using a subjective measure of workload) showed that the two levels of auditory load had the expected effects. It is worth noting that in their experiment, an aural alert (used to initiate a check list task) was just as effective as a tactile alert even in varying conditions of auditory load. Oskarsson et al. (2008) used a very similar auditory secondary task where participants were required to listen for radio calls which were composed of colour and number combinations. When a radio call occurred, the participant would acknowledge the call sign by pressing the corresponding button on a touch screen. Both experiments used the proportion of correctly answered radio calls as a dependent measure, while the Oskarsson et al. (2008) also measured response time.

The final type of secondary task used was a check list completion task. Check list tasks were used in Calhoun et al. (2003) and its follow-up experiment, Calhoun et al. (2004). In both experiments, participants were asked to monitor alerts for different priorities presented through different combinations of visual, aural, and tactile cues. Different combinations of cues would represent different types of warnings. Each warning had a specific check list of tasks associated with it. Each check list started off with an acknowledgement that the participant had detected the alert, and the number of tasks required in the check list was used to vary the difficulty of the task. The dependent measure used was proportion correct and time until detection.

7.3 Potential Experiment Ideas

The following section describes potential experiment ideas that could be followed based on the results and topics discussed in this literature review.

7.3.1 Derive Multimodal Requirements from EID

Synopsis: Using the EID framework, analytically determine which requirements might be suitable for tactile or auditory display and for what reason. This approach requires further elaboration of the EID framework which does not typically specify the mode of display of information.

Benefits: This would clearly be a novel contribution that has theoretical and practical benefits. The EID framework would be improved by providing more information on how to implement the requirements that it suggests. From a practical standpoint, this could provide an analytical approach to determining which variables should be displayed in the tactile or auditory form.

Requirements: An abstraction hierarchy of an aircraft was developed at the Advanced Interface Design Lab (AIDL) several years ago. This covers the basic *work domain analysis* of any aircraft and can be used to guide this process. This reduces much of the work for this option and allows the work to concentrate specifically on the modality of the requirements.

7.3.2 Explore Ecological Tactile and Auditory Displays

Synopsis: There is some evidence that tactile and auditory display may be quite effective for displays of ambient, system-health type information. This connects well with information that is typically at the higher levels of the abstraction hierarchy. The auditory display is not as novel as this has been looked at by Sanderson but the tactile display would be quite a novel contribution. This approach requires experimentation to determine whether a tactile or auditory display of higher level information is useful.

Benefits: This project would connect well with Option #1, providing some experimental evidence of whether a tactile or auditory ecological display would be useful. It provides greater focus than Option #1 as it looks mostly at the display of higher level information.

Requirements: From an Abstraction Hierarchy (AH) of an aircraft, abstract functional variables should be identified. One or two of these should be further selected for tactile or auditory display. Experimentally the "ecological" multimodal display should be tested to see whether the ecological version improves understanding of situation. Potential variables for consideration might be: engine health, groundspeed (sum of windspeed and airspeed), and time to decision point.

Note: Higher level functional information is not likely available on the GCS interface. This creates the situation where one interface has more information than the other, which can confound results.

Table 10: Possible Experiment Designs to Explore Ecological Auditory and Tactile Displays.

Possible Experimental Designs	**Conditions**	**Notes**
Design A: Effect of multimodal EID	Condition 1: Baseline	Likely different information in condition 2. This creates a confound. These confounds do happen (note original duress studies all had the same
	Condition 2: Baseline + mm EID	

		confound)
Design B: Effect of EID in visual vs. multimodal modalities	Condition 1: Baseline Condition 2: Baseline +visual EID Condition 3: Baseline + mmEID	This design eliminates the above confound but requires further work in designing the visual EID and developing a third interface. The experimental design with 3 conditions also adds complexity in interpreting results.
Design C: Simplified version of Design B	Condition 1: Baseline +visual EID Condition 2: Baseline + mmEID	Essentially the same design as above but does not run the straight baseline condition. The multimodal question still gets explored but the benefit of the EID information cannot be determined.
Design D: Effect of EID in auditory vs. tactile modalities.	Condition 1: Baseline Condition 2: Baseline + auditory Condition 3: Baseline + tactile	The objective of this design is to tease out whether the modality of the display is an influencing factor on performance. Auditory and tactile displays must be carefully constructed to ensure they convey the same information. There is a risk that there is no effect at all, so a pilot study would be recommended in this case.
Design E: Simplified Design D	Condition 1: Baseline + auditory Condition 2: Baseline + tactile	The same design but without running the straight baseline condition.
Design F: Effect of EID in auditory, tactile and redundant modalities.	Condition1: Baseline Condition 2: Baseline + auditory Condition 3: Baseline + tactile Condition 4: Baseline + auditory + tactile	This design could show if there is a beneficial effect of added redundancy through modalities. Could also be run without condition 1 for some experiment efficiency. The disadvantage is the large number of conditions. Pilot testing is highly recommended.

7.3.3 Tacton, Tactile Display Design

Synopsis: There is very little known about tactile display in comparison with visual display design. Across a display space as flexible as tactile vest, many options arise from a tacton design, tactile icon design, and *tactification*.

Benefits: This project provides clear guidance on tactile display design. It is also necessary in order to design an effective tactile display for the full multimodal display.

Requirements: A small experiment (not involving the simulation) to determine how quickly people can understand various tactile designs. The task should be short so many trials can be run quickly, 20 participants or less.

Table 11: Possible Experimental Designs to Investigate Tacton and Tactile Display Design.

Possible Experimental Designs	**Conditions**	**Notes**
Design A: Comparison of tactile forms of reference	Condition 1: Iconic tacton. Tacton "feels like" something related in the real world	In some cases, the iconic tacton may not really be possible to develop.
	Condition 2: Propositional tacton. Tacton acts a symbol for something in the real world.	
	Condition 3: Analogical tacton. Tacton presents a mapping for something in the real world.	
Design B: Simplified Design A	Condition 1: Propositional tacton. Tacton acts a symbol for something in the real world.	Eliminates the iconic tacton if that does not seem to be feasible.
	Condition 2: Analogical tacton. Tacton presents a mapping for something in the real world.	
Design C: Tactification	Follow designs A or B with a tactification condition	True tactification as in a signal being directly produced on the tactors may not be technically feasible for us. It may also not be comfortable for the operator.
Design D: Comparison of tacton designs	Condition 1: Design 1.	Given various feasible design alternatives, this would be a simple test to isolate the most
	Condition 2: Design 2.	
	Condition 3: Design 3.	

		promising mapping. This could be combined with Design A or B.

7.3.4 Cue Prioritization Study

Synopsis: There is clear evidence that some landing scenarios are very challenging and the UAV operator is deprived of information simply by being located on the ground and not in the aircraft. However, applying multimodal technology effectively requires acknowledging that information from different modalities may have cross-modal interactions. It would be worthwhile to investigate whether multimodal information can improve the performance of UAV operators when they are presented with information in different modalities and the most salient visual information is not the most important information in managing the landing scenario. As an example, the particular GCS interface in this project has a couple very salient visual features specific to landing, those being altitude and distance to the landing location. This information is conveyed in large bright green bars that also include the decision points for the abort decision. A second highly salient feature is the glide slope indicator. This indicator quickly tells the operator if they have deviated from the ideal glide slope (though note at this time, the abort decision box is not implemented on the interface).

However, there are many other important variables to consider in landing the aircraft and, while visible, these variables are displayed in a low salient, text based format. Quite reasonably one or more of these variables (for example wind speed, or ground speed) could be supplemented through a redundant tactile display on the tactor vest or through auditory information. We would hypothesize that the redundant information would now become more salient and result in better performance. As a concern however, presenting the redundant information could be distracting or create additional channel loading, particularly in scenarios that did not require it. This interaction of channel loading with redundant display potentially reordering cue priorities creates an interesting research question.

Benefits: Practical information on how to apply multimodal technology. Theoretical insight into cue dominance and channel loading.

Requirements: To study channel loading a secondary task needs to be used. This needs to be implemented in time for the baseline study in order to capture baseline performance. There are a few options here: a modality specific secondary task would test loading in the visual or tactile modalities, but could also arguably create load. An auditory secondary task presents opportunity for some realism (a task with auditory air traffic chatter or chatter within the unit might be a good choice), and loading that does not interfere with the channels but would only be informative in terms of general cognitive load, not particularly modal loading.

Note: a cue conflict study is not a realistic alternative as it would be unlikely that that the visual and tactile information would differ due to the design of these systems (i.e. a single processor that outputs to the two interface methods). If there were a realistic argument for two independently calculated variables that could differ, this could be easily adjusted to a cue conflict study.

Table 12: Possible Experiment Designs to Explore Cue Prioritization.

Possible Experimental Designs	**Conditions**	**Notes**
Design A: Auditory cue prioritization, straight performance	Condition 1: Baseline	Without a secondary task, there is a risk that performance results are the same in each condition. This can come about as a result that the task itself is not usually very difficult.
	Condition 2: Baseline + auditory cues	
Design B: Auditory cue prioritization, channel loading	Same conditions as A but using an auditory secondary task.	There is a reasonable chance that if performance differences do not show on the primary task, they may show on the secondary task. This would confirm the suspicion of channel loading. The issue with this design though is how these results would be interpreted. If performance is the same, but the auditory channel is loaded, would that not suggest that adding the auditory cues makes things worse?
Design C: Auditory cue prioritization, visual channel loading	Same conditions as A but with a visual secondary task.	The objective of this experiment is to see whether adding the auditory cues reduces visual channel loading.
Design D: Tactile cue prioritization series	Same models as A, B, and C but with tactile cuing.	
Design E: Multimodal cueing series	Same models as A, B, and C but with multimodal cuing.	In this design, all modalities are working. There could be some complex effects. For one, assuming the visual and auditory variables are the same, redundancy should either improve performance or reduce loading. However, if the auditory and tactile variables are not redundant, complex cross-modal effects

		may be seen that may be very interesting, but nearly impossible to interpret. This design is not recommended at this stage of our knowledge of how these displays work.

7.3.5 Warning Study

Synopsis: One way of using multimodal display may be for alerting. Currently auditory displays are regularly used in this way and are well understood. Tactile displays may present a way to alert operators in situations where auditory loading is high.

Benefits: An understanding of how to use tactile displays to capture attention. As auditory displays are relatively well understood, the greatest benefit lies in understanding the tactile displays.

Requirements: An auditory secondary task would be useful in order to understand whether tactile displays present an opportunity in dense auditory spaces.

Table 13: Possible Experimental Designs to Explore Warnings

Possible Experimental Designs	**Conditions**	**Notes**
Design A: Tactile as a redundant supplement to auditory and visual warnings	Condition 1: Auditory + visual warnings	If the task is not complex enough, results may not be seen in this experiment as all participants will be able to respond quickly regardless of modality.
	Condition 2: Auditory + visual + tactile warning	
Design B: Tactile warnings at different levels of auditory loading	Conditions: Auditory + visual + tactile warnings tested across a range of auditory chatter levels (e.g. low medium high). An auditory secondary task could be added to increase loading further.	The secondary task may not be needed if the chatter levels are high enough.

7.3.6 Training Study

Synopsis: There is strong motivation to create a “UAV operator” that does not have the flight education and in-flight experience of a pilot. An operator trained this way could be trained quickly and much less expensively than a traditional pilot. The risks to this approach are that operators trained this way may not properly understand the dynamics of flight and the aircraft and as a result make more and more costly errors.

Since the landing task available in the simulation is relatively constrained, with easily understandable rules, it should be possible to train novices to land the simulated UAV within a reasonable amount of training time. This study should compare two sets of novices using different training regimes to determine the effect of training on novice performance. Comparison with a baseline of "experts", UAV operators with in-flight experience would allow for adequate consideration of whether the novices could potentially be adequately trained to be competent UAV operators without in-flight experience.

Benefits: Practical guidance on training UAV operators.

Requirements: A baseline study with pilots should be run to have a benchmark of pilot performance. A stronger understanding of UAV operator training approaches would need to be obtained. Ideally, there should be some theoretical grounding to the training (e.g. experience with critical incident scenarios vs. book training).

Note: the tactile condition is not necessarily needed in this option.

Design: 2x2 with novices/experts as one, training method 1/training method 2 as the other.

7.3.7 Automation Study

Synopsis: Understanding unreliable automation continues to be a problem in many areas and UAV operation is highly automated. This study would use the tactile modality to improve information when automation is not reliable.

Requirements: The autoland automation must be accessible to manipulate its reliability. This is currently not possible, or at least under debate. One would have to consider whether the baseline visual interface required modification to indicate the reliability of the automation.

Note: not recommended at this time. Potentially an option if it turns out the reliability can be influenced.

7.3.8 Intelligent Adaptive Interface Study

Synopsis: Multimodal technology presents the opportunity to move or reinforce visual information through other modalities. With contextual capture of either the workload or the physical state of the operator, context or operator state could be used as triggers to adapt either the modality of the information presentation. Two clear options present themselves – first key variables could be supplemented with multimodal display based on either high workload or high operator stress. Alternatively warning signals could be displayed in non-visual modalities based on various triggers.

Benefits: New information on how to use multimodal displays in an adaptive interface context.

Requirements: Appropriate triggers for adaptation must be defined. Some potential candidates would be:

1. Operator workload: This could be measured by scenario context, or by operator action frequency such as interactions with the UAV system. One issue with the latter approach is that the high level of automation does not make this high workload task.
2. Operator attention: The eye-tracking system could be used to monitor operator visual attention and provide multimodal alerts if the operator is not looking at the visual display.
3. Operator auditory loading: The degree of auditory chatter could be measured and used to trigger supplemental multimodal alerting in cases where the auditory channel is overloaded. Note that experiment proposal 4 could begin to provide clues to when that channel is overloaded.
4. Operator stress: Heart or Galvanic skin response could be used as an indicator of stress to trigger the adaptation of the display. Further research would be needed to confirm appropriate trigger levels. Note: with the high degree of automation in the current simulation stress levels may remain quite low.

8 Conclusion

In this report we have completed a literature review of topics that can support research about multimodal interfaces with a focus on different methods of multimodal information presentation, and issues with perception of the multisensory information by human observers. Multimodal interfaces present an exciting and relatively untapped method for improving the flow of information from the system to the user.

Currently, there are very few fully developed design methodologies for exploiting multimodal interfaces. Much of the research has been focused on either designing for specific modalities, or on the perception of stimuli in different modalities. We explored the use of one possible design methodology, EID, as examined how it has been used for non-visual interface design in the past, and what issues still exist with using this methodology to design multimodal interfaces. EID appears to be a method that can be adapted for use with multimodal interfaces. However, a focus on tasks and attention requirements must be added to EID to effectively assist with multimodal interface design.

In addition, interface designers must have a solid grasp of how users will perceive information in each modality that they design for, no matter what design methodology that use. We reviewed both tactile and auditory perception within this report. Tactile perception is still a relatively new field of research, and the use of tactile displays in real-world applications is still limited. Much of the focus is still focused on understanding basic perceptual issues. We noticed that very few of these experiments contained detailed descriptions of how they selected the semantic mapping and information coding methods that were used. This lack of a systematic design methodology can be supported with the use of methods such as EID. Auditory perception, as a field, is much further developed, and research has turned more towards understanding how concepts and data can be coded and mapped onto auditory stimuli and presented to users. There is evidence that human observers have preconceptions of how different types of data “should” sound, and therefore there are intuitive methods for displaying information in the auditory modality that improve memorability and response time. However, it is important to note that different user populations may have different preconceptions of how data should be coded into auditory characteristics.

The use of multiple modalities also raises questions of how the user’s attention will be directed by the different channels of information. In a traditional single display visual interface, designers can assume that the locus of the user’s attention is on the display. However, as more displays are added, it becomes much more difficult to estimate what the user is focused on. This problem is compounded when multiple modalities are used, since the possible display space is vastly increased. We presented different possible models of how the human attention system may work, but this is still no conclusive evidence that supports any one model. However, research has shown that there are strong ties between attention direction in different modalities, and the degree to which the modalities interact may be a function of the type of task being preformed. As such, preliminary multimodal interface designs cannot assume that information in each modality can be interpreted independently of other modalities.

As multimodal interfaces become more popular, the possibility of designing even more complex adaptive multimodal interfaces becomes more likely. We reviewed guidelines for the design of IAI systems, and examined how these could be applied to future adaptive multimodal interfaces. Since we still lack of full understanding of what multimodal interfaces will entail, it may still be too early to identify ways these interfaces can be made adaptive. However, there exists great potential with the intersection of these two ideas.

Finally, we presented information related to autolanding problems with UAVs. We examined situations where a multimodal interface may be appropriate due to off-nominal events or complex situations. We also examined methodologies used in similar UAV, autoland, or fault detection studies. This was done to provide insight into possible methods of evaluating future multimodal interface designs. We concluded by providing a list of possible experiment ideas that will explore the use and design of multimodal interfaces for the UAV autolanding scenario.

9 References

Anderson, J. E., & Sanderson, P. (2004). Designing sonifications for effective attentional control in complex work domains. In *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 1818-1822). Santa Monica, CA: Human Factors and Ergonomics Society.

Anderson, J. E., & Sanderson, P. (2009). Sonification design for complex work domains: dimensions and distractors. *Journal of experimental psychology. Applied*, *15*(3), 183-98.

Aretz, D., Andre, T., Self, B., & Brenaman, C. (2006). Effect of tactile feedback on unmanned aerial vehicle landings. In *Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC).*

Arons, B. (1992). A Review of The Cocktail Party Effect. *Journal of the American Voice I/O Society*, *12*, 35-50.

Arrabito, G. R., Mondor, T., & Kent, K. (2004). Judging the urgency of non-verbal auditory alarms: a case study. *Ergonomics, 47*(8), 821-840.

Beierholm, U. R., Kording, K. P., Shams, L., & Ma, W. J. (2007). Comparing Bayesian models for multisensory cue combination without mandatory integration. In *Proceedings of the 21st Annual Conference on Neural Information Processing Systems (NIPS 2007).*

Belz, S. M., Robinson, G. S., & Casali, J. G. (1999). A new class of auditory warning signals for complex systems: Auditory icons. *Human Factors*, *41*, 608-618.

Besnard, D., Greathead, G., and Baxter, G., (2004) International Journal of Human-Computer Studies. *When mental models go wrong. Co-occurrences in dynamic, critical systems,* Vol. 60, pp. 117-128.

Billings, C.E. (1991). *Human-centered aircraft automation: A concept and guidelines* (NASA TM 103885). Moffett Field, CA: NASA Ames Research Center.

Billings, C.E. (1996). *Human-centered aviation automation: principles and guidelines* (NASA TM 110381). Moffett Field, CA: NASA Ames Research Center.

Bliss, J. C., Katcher, M. H., Rogers, C. H., & Shepard, R. P. (1970). Optical-to-tactile image conversion for the blind. *IEEE Transactions on Man–Machine Systems*, 11(4), 58-64.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: The MIT Press.

Brewster, S. A., & Brown, L. M. (2004). Tactons: Structured tactile messages for non-visual information display. In *Proceedings of the 5th Australasian User Interface Conference* (pp. 15-23). Sydney, Australia: Australian Computer Society.

Brill, J. C., Mouloua, M., Gilson, R. D., Rinalducci, E. J., & Kennedy, R. S. (2008). Effects of secondary loading task modality on attentional reserve capacity. In *Proceedings of the 52nd Annual Meeting of the Human Factors and Ergonomics Society* (pp. 1219-1223). Santa Monica, CA: Human Factors and Ergonomics Society.

Brown, L. M., Brewster, S. A., & Purchase, H. C. (2005). A first investigation into the effectiveness of tactons. In *Proceedings of the First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems* (pp. 167-176). Washington, DC: IEEE Computer Society.

Brown, L. M., Brewster, S.A., Purchase, H. C. (2006a). Multidimensional tactons for non-visual information presentation in mobile devices. In *Proceedings of the 8th Conference on Human-computer Interaction with Mobile Devices and Services* (pp. 231-238). New York, NY: Association for Computing Machinery.

Bruseberg, A., & Johnson, P. (2004). *Should Computers Function as Collaborators?*. In Proceedings of HCI-Aero 2004 held in Toulouse, France September 29, 2004 to 1 October 2004

Burns, C. M. (2000). Putting it all together: Improving display integration in ecological displays. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 42(2), 226-241.

Burns, C. M., & Hajdukiewicz, J. (2004). Ecological Interface Design. Boca Raton, FL: CRC Press.

Calhoun, G., Draper, M., Ruff, H., Fontejon, J., & Guilfoos, B. (2003). Evaluation of tactile alerts for control station operation. In *Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 2118-2122). Santa Monica, CA: Human Factors and Ergonomics Society.

Calhoun, G., Fontejon, J., Draper, M., Ruff, H., & Guilfoos, B. (2004). Tactile versus aural redundant alert cues for UAV control applications. In *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 137-141). Santa Monica, CA: Human Factors and Ergonomics Society.

Cheung, B., van Erp, J. B. F., & Cholewiak, R. W. (2008). Anatomical,neurophysiological and perceptual issues of tactile perception. In J. B. F. van Erp & B. P. Self (Eds.), *Tactile displays for orientation, navigation and communication in air, sea and land environments (Report No. TR-HFM-122, p. 2-1 - 2-18).* Neuilly-sur-Seine Cedex, France: North Atlantic Treaty Organization, Research and Technology Organization.

Cholewiak, R. W., Collins, A. A., & Brill, C. J. (2001). Spatial factors in vibrotactile pattern perception. In C. Baber, M. Faint, S. Wall, & A. M. Wing (Eds.), *Eurohaptics 2001: Conference Proceedings*.

Cholewiak, R. W., & Collins, A. A. (2000). The generation of vibrotactile patterns on a linear array: Influences of body site, time, and presentation mode. *Perception & Psychophysics*, 62(6), 1220-1235.

Cholewiak, R. W., & Collins, A. A. (2003). Vibrotactile localization on the arm: Effects of place, space, and age. *Perception & Psychophysics*, 65(7), 1058-1077.

Cholewiak, R. W., & Collins, A.A. (1991). Sensory and physiological bases of touch. In M. A. Heller,& W. Schiff (Eds.), *The Psychology of Touch*, 23-60. Hillsdale, NJ: Erlbaum.

Cholewiak, R. W., Brill, J. C., & Schwab, A. (2004). Vibrotactile localization on the abdomen: Effects of place and space. *Perception & Psychophysics*, 66(6), 970-987.

Chouvardas, V. G., Miliou, A. N., & Hatalis, M. K. (2005). Tactile displays: A short overview and recent developments. In *Proceedings of the 5th International Conference on Technology and Automation* (pp. 246-251).

Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics*, 16(2), 409-412.

Craig, J. C. (1972). Difference threshold for intensity of tactile stimuli. *Perception & Psychophysics*, 11(2), 150-152.

Craig, J. C. (1974). Vibrotactile difference thresholds for intensity and the effect of a masking stimulus. *Perception & Psychophysics*, 15(1), 123-127.

Craig, J. C., & Evans, P. M. (1987). Vibrotactile masking and the persistence of tactual features. *Perception & Psychophysics*, 42(4), 309-317.

Croft, B. (1995). What Do People Want from Information Retrieval? *D-Lib Magazine*. Retrieved from http://www.dlib.org/dlib/november95/11croft.html.

Davies, T. C., Burns, C. M., & Pinder, S. D. (2007). Testing a novel auditory interface display to enable visually impaired travelers to use sonar mobility devices effectively. In *Proceedings of the 51st Annual Meeting of the Human Factors and Ergonomics Society* (pp. 278-282). Santa Monica, CA: Human Factors and Ergonomics Society.

Donmez, B., Graham, H., & Cummings, M. (2008). *Assessing the Impact of Haptic Peripheral Displays for UAV Operators* (Report No. HAL2008-02). Cambridge, MA: MIT Humans and Automation Laboratory. Retrieved from http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA479798&Location=U2&doc=GetTRDoc.pdf

DRDC. (2009). *Statement of Work for Advanced Interface UAV Human Factors R&D support to DRDC Toronto*. Toronto, ON: National Defence and Canadian Forces.

Edworthy, J., & Hellier, E. (2006). Alarms and human behaviour: Implications for medical alarms. *British Journal of Anaesthesia*, 97 (1), 12-17.

Endsley M, Kaber D (1997). *Out-of-the-Loop Performance Problems and the Use of Intermediate Levels of Automation for Improved Control System Functioning and Safety*. Process Safety Progress, 126-130.

Farah, M. J., Wong, A. B., Monheit, M. A., & Morrowt, L. A. (1989). Parietal lobe mechanism of spatial attention: modality-specific or supramodal?. *Neuropsychologia*, 27(4), 461-470.

Franconeri, S. L., Hollingworth, A., & Simons, D. J. (2005). Do new objects capture attention?. *Psychological Science*, 16(4), 275-281.

Gallace, A., & Spence, C. (2009). The cognitive and neural correlates of tactile memory. *Psychological Bulletin*, 135(3), 380-406.

Gallace, A., Tan, H. Z., & Spence, C. (2007). The body surface as a communication system: The state of the art after 50 years. *Teleoperators and Virtual Environments*, 16(6), 657-676.

Galvin, K. L., Mavrias, G., Moore, A., Cowan, R. S., Blamey, P. J., & Clark, G. M. (1999). Acomparison of Tactaid II+ and Tactaid 7 use by adults with a profound hearing impairment. *Ear and Hearing*, 20, 471-482.

Gawron, V.J. (2008). *Human Performance, Workload, and Situational Awareness Measures Handbook*. Boca Raton, FL: CRC Press.

Gescheider, G., Bolanowski, S., & Verrillo, R. (1989). Vibrotactile masking: effects of stimulus onset asynchrony and stimulus frequency. The Journal of the Acoustical Society of America, 85(5), 2059-64. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2732386.

Goldstein, I. L., & Dorfman, P. W. (1978). Speed and load stress as determinants of performance in a time sharing task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 20(5), 603-609.

Greenspan, J. D., & Bolanowski, S.J. (1996). The psychophysics of tactile perception and its peripheral physiological basis. In L. Kurger (Ed.), *Pain and Touch*, 2, 25-103.

Haas, E., & Edworthy, J. (1996). Designing urgency into auditory warnings using pitch, speed and loudness. *Computing \& Control Engineering Journal*, *7*(1), 193. Retrieved from http://link.aip.org/link/?CCEJEL/7/193/1.

Hager, G., Kriegman, D., & Morse, A. (1998). Research issues in vision and control. *Lecture Notes in Control and Information Sciences*, 237, 1-10.

Hameed, S., & Sarter, N. (2009). Context-sensitive information presentation: Integrating adaptive and adaptable approaches to display design. In *Proceedings of the 53rd Annual Meeting of the Human Factors and Ergonomics Society* (pp. 1694-1698). Santa Monica, CA: Human Factors and Ergonomics Society.

Hansen, J. P. (1995). An experimental investigation of configural, digital, and temporal information on process displays. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(3), 539-552. Santa Monica, CA: Human Factors and Ergonomics Society.

Hansman, R.J. (2001). *Complexity of Aircraft Automation – A precursor for concerns in Human-Automation Systems.* National Forum.

Healey, C. G., Booth, K. S., & Enns, J. T. (1996). High-speed visual estimation using preattentive processing. *ACM Transactions on Computer-Human Interaction*, 3(2), 107-135.

Hellier, E., & Edworthy, J. (1999). On using psychophysical techniques to achieve urgency mapping in auditory warnings. *Applied Ergonomics*, *30*(2), 167–171.

Helbig, H. B., & Ernst, M. O. (2007). Knowledge about a common source can promote visual-haptic integration. *Perception*, 36(10), 1523-1533.

Hillstrom, A. P., Shapiro, K., & Spence, C. (2002). Attentional and perceptual limitations in processing sequentially presented vibrotactile targets. *Perception & Psychophysics*, 64(7), 1068-1082.

Ho, C., & Spence, C. (2008). *The Multisensory Driver: Implications for Ergonomic Car Interface Design*. Aldershot, England: Ashgate Publishing Limited.

Ho, C., & Spence, C. (2009). Using peripersonal warning signals to orient a driver's gaze. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 51(4), 539-556.

Ho, C., Nikolic, M. I., & Sarter, N. B. (2001). Multimodal information presentation in support of timesharing and effective interruption management. In *Proceedings of the 20th Digital Avionics Systems Conference* (pp.5D2/1-5D2/8).

Hoggan, E., & Brewster, S. (2007). Designing Audio and Tactile Crossmodal Icons for Mobile Devices. *Design*, 162-169.

Hou, M., Gauthier, M. S., & Banbury, S. (2007a). Development of a generic design framework for intelligent adaptive systems. *Human-Computer Interaction*, 313-320.

Hou, M., Kobierski, R. D., & Brown, M. (2007b). Intelligent adaptive interfaces for the control of multiple UAVs. *Journal of Cognitive Engineering and Decision Making*, 1(3), 327-362.

Huffman, K. (2007). *Psychology in Action* (Eighth Edition.). United States of America: Hermitage Publishing Services.

Jones, L. A., & Sarter, N. B. (2008). Tactile displays: Guidance for their design and application. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(1), 90-111.

Jones, L. A., Lockyer, B., & Piateski, E. (2006). Tactile display and vibrotactile pattern recognition on the torso. *Advanced Robotics*, 20(12), 1359-1374.

Kaaresoja, T., & Linjama, J. (2005). Perception of short tactile pulses generated by a vibration motor in a mobile phone. In *Proceedings of the First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems* (pp. 471-472). Los Alamitos, CA: IEEE Computer Society.

Kandel, E. R., Schwartz, J. H., & Jessel, T. M. (1991). *Principles of Neural Science*. New York, NY: Elsevier Science Publishing Co.

Karnath, H. O., Schenkel, P., & Fischer, B. (1991). Trunk orientation as the determining factor of the "contralateral" deficit in the neglect syndrome and as the physical anchor of the internal representation of body orientation in space. *Brain*, 114(4), 1997-2014.

Kirman, J. (1974). Tactile apparent movement: The effects of interstimulus onset interval and stimulus duration. *Perception &Psychophysics*, 15(1), 1-6.

Kitagawa, N. (2006). Audiotactile multisensory interactions in human. *Japanese Psychological Research*, 48(3), 158-173.

Kitagawa, N., Zampini, M., & Spence, C. (2005). Audiotactile interactions in near and far space. *Experimental Brain Research*, 166, 528-537.

Koppen, C. M., & Spence, C. (2007c). Spatial coincidence modulates the Colavita visual dominance effect. *Neuroscience Letters*, 417(2), 107-111.

Koppen, C., & Spence, C. (2007a). Assessing the role of stimulus probability on the Colavita visual dominance effect. *Neuroscience Letters*, 418(3), 266-271.

Koppen, C., & Spence, C. (2007b). Audiovisual asynchrony modulates the Colavita visual dominance effect. *Brain Research*, 1186(1), 224-232.

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS ONE*, 2(9), e943.

Kramer, L., & Busquets, A. (2000). Comparison of Pilots' Situational Awareness While Monitoring Autoland Approaches Using Conventional and Advanced Flight Display Formats. Hampton, Virginia, USA.

Langford, N., Hall, R. J., & Monty, R. A. (1973). Cutaneous perception of a track produced by a moving point across the skin. *Journal of Experimental Psychology*, 97(1), 59-63.

Lederman, S. J., Klatzky, R. L. (2009). Haptic perception: A tutorial. *Attention, Perception & Psychophysics*, 71(7), 1439-1459.

Lederman, S. J., Thorne, G., & Jones, B. (1986). Perception of texture by vision and touch: Multidimensionality and intersensory integration. *Journal of Experimental Psychology: Human Perception and Performance*, 12(2), 169-180.

Lee, J., Stoner, H., & Marshall, D. (2004). Enhancing interaction with the driving ecology through haptic interfaces. *IEEE International Conference on Systems, Man and Cybernetics*, 1, 841-846.

Ley, I., Haggard, P., & Yarrow, K. (2009). Optimal integration of auditory and vibrotactile information for judgments of temporal order. *Journal of Experimental Psychology: Human Perception and Performance*, 35(4), 1005-1019.

Lindeman, R. W., & Yanagida, Y. (2003). Empirical studies for effective near-field haptics in virtual environments. In *Proceedings of the 2003 IEEE Virtual Reality Conference* (pp. 287-288). Los Alamitos, CA: IEEE Computer Society.

Lukas, S., Philipp, A. M., & Koch, I. (2010). Switching attention between modalities: Further evidence for visual dominance. *Psychological Research*, 74(3), 255-267.

Ma, W. J., & Pouget, A. (2008). Linking neurons to behavior in multisensory perception: A computational review. *Brain Research*, 1242, 4-12.

Maat, L., & Pantic, M. (2006). Gaze-X: Adaptive affective multimodal interface for single-user office scenarios. In *Proceedings of the 8th International Conference on Multimodal* Interfaces (pp. 171-178). New York, NY: Association for Computing Machinery.

McDonald, J. J., Teder-sälejärvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature*, 407, 906-908.

McNeer, R. R., Bohorquez, J., Özdamar, Ö., Varon, A. J., & Barach, P. (2007). A new paradigm for the design of audible alarms that convey urgency information. *Journal of Clinical Monitoring and Computing*, 21(6), 353-363.

Meyer, B., Yakemovic, K., & Harris, M. (1993). Issues in practical application of an adaptive interface. In *Proceedings of the 1st International Conference on Intelligent User Interfaces* (pp. 251-254). New York, NY: Association for Computing Machinery.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limitations on our capacity for processing information. *Psychological Review*, 63, 81-97.

Mitchell P.J., Cummings M.L., (2005). *Management of multiple dynamic human supervisory control tasks.* MIT Press.

Moray, N. (1981). The role of attention in the detection of errors and the diagnosis of failures in man-machine systems. In J. Rasmussen,& W. B. Rouse (Eds.), *Human Detection and Diagnosis of System Failures* (pp.185-198). New York, NY: Plenum Press.

Moray, N., & Inagaki, T. (2000). Attention and complacency. *Theoretical Issues in Ergonomics Science*, 1(4), 354-365.

Nafe, J. P., & Wagoner, K. S. (1941). The nature of pressure adaptation. *Journal of General Psychology*, 25, 323-351.

Neuhoff, J.G. (Ed.). (2004). *Ecological Psychoacoustics*. San Diego, CA: Elsevier Academic Press.

Occelli, V., O'Brien, J. H., Spence, C., & Zampini, M. (2010). Assessing the audiotactile Colavita effect in near and rear space. *Experimental brain research.* Retrieved May 18, 2010, from http://www.springerlink.com/content/ll77m9u23k450230/

Orlady H, Orlady L (1999). *Human Factors in Multi-Crew Flight Operations* Ashgate Press.

Oskarsson, P., Eriksson, L., Lif, P., Lindahl, B., & Hedström, J. (2008). Multimodal threat cueing in simulated combat vehicle. In *Proceedings of the 52nd Annual Meeting of the Human Factors and Ergonomics Society* (pp. 1287-1291). Santa Monica, CA: Human Factors and Ergonomics Society.

Pattyn, N., Neyt, X., Henderickx, D., & Soetens, E. (2008). Psychophysiological investigation of vigilance decrement: Boredom or cognitive fatigue?. *Physiology & Behavior*, 93(1-2), 369-378.

Pentland, A., & Roy, D. (1998). Multimodal adaptive interfaces. In *Papers from the 1998 AAAI Spring Symposium on Intelligent Environments* (pp. 115-122). Menlo Park, CA: Association for the Advancement of Artificial Intelligence.

Priplata, A. A., Niemi, J. B., Harry, J. D., Lipsitz, L. A., & Collins, J. J. (2003). Vibrating insoles and balance control in elderly people. *Lancet*, 362(9390), 1123-1124.

Rees, G., Frith, C., & Lavie, N. (2001). Processing of irrelevant visual motion during performance of an auditory attention task. *Neuropsychologia*, 39, 937-949.

Reeves, L. M., Lai, J., Larson, J. A., Oviatt, S., Balaji, T. S., Buisine, S . . . Wang, Q. Y. (2004, January). Guidelines for multimodal user interface design. *Communications of the ACM*, 47(1), 57-59.

Roach, N. W., Heron, J., & McGraw, P. V. (2006, September). Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. *Proceedings of the Royal Society Biological Sciences*, 273(1598), 2159-2168.

Rothenberg, M., Verrillo, R. T., Zahorian, S. A., Brachman, M. L., & Bolanowski, S. J. (1977). Vibrotactile frequency for encoding a speech parameter. *Journal of the Acoustical Society of America*, 62(4), 1003-1012.

Rupert, A. H. (2000, March-April). An instrumentation solution for reducing spatial disorientation mishaps. *IEEE Engineering in Medicine and Biology Magazine*, 19(2), 71-80.

Sanders, M. S. & McCormick, E. J. (1993). *Human Factors In Engineering and Design* (7th ed.). Columbus, OH: McGraw-Hill.

Sanderson, P. M., & Watson, M. O. (2005). From information content to auditory display with ecological interface design: Prospects and challenges. In *Proceedings of the 49th Meeting of the Human Factors and Ergonomics Society* (pp. 259-263). Santa Monica, CA: Human Factors and Ergonomics Society.

Sanderson, P., Anderson, J., & Watson, M. (2000). Extending ecological interface design to auditory displays. In *Proceedings of the 2000 Annual Conference of the Computer-Human Interaction Special Interest Group (CHISIG) of the Ergonomics Society of Australia (OzCHI2000)* (pp. 259-266).

Sanderson, P., Wee, A., Seah, E., Lacherez, P. (2006). Auditory alarms, medical standards, and urgency. In *Proceedings of the 12th International Conference on Auditory Display*, (pp. 24-27).

Santangelo, V., Fagioli, S., & Macaluso, E. (2010). The costs of monitoring simultaneously two sensory modalities decrease when dividing attention in space. *NeuroImage*, 49(3), 2717-2727.

Sarter, N. B. (2006). Multimodal information presentation: Design guidance and research challenges. *International Journal of Industrial Ergonomics*, 36(5), 439-445.

Sarter, N. B. (2007). Multiple resource theory as a basis for mul- timodal interface design: Success stories and qualifications. In A. Kramer, D. Wiegmann, & A. Kirlik (Eds.), *Attention: From Theory to Practice* (pp. 187-195). New York, NY: Oxford University Press.

Sarter, N.B., & Woods, D.D. (1992). Pilot interaction with cockpit automation: Operational experiences with the Flight Management System. *International Journal of Aviation Psychology 2*(4), 303-321.

Sarter, N.B., & Woods, D.D. (1994). Pilot interaction with cockpit automation II : An experimental study of pilot's model and awareness of the Flight Management System. *International Journal of Aviation Psychology 4*(1), 1-28.

Sarter, N.B., Woods, D.D. (1995). 'How in the world did we ever get into that mode?' Mode error awareness in supervisory control. *Human Factors 31*(1), 5-19.

Schneider-Hufschmidt, M., Groh, L., Perrin, P., Hine, N., & Furner, S. (2003). Human Factors guidelines for multimodal interaction, communication and navigation. In *Proceedings of the 19th International Symposium on Human Factors in Telecommunication*.

Scott, J. J., & Gray, R. (2008). A comparison of tactile, visual, and auditory warnings for rear-end collision prevention in simulated driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(2), 264-275.

Sheridan, T. (1992). *Telerobotics, automation, and human supervisory control*. Cambridge, Massachusetts: The MIT Press.

Sherrick, C. A., Cholewiak, R.W.,& Collins, A. A. (1990). The localization of low- and high-frequency vibrotactile stimuli. *The Journal of the Acoustical Society of America*, 88(1), 169-179.

Sherrick, C. E. (1985). A scale for rate of tactual vibration. *The Journal of the Acoustical Society of America*, 78(1), 78-83.

Sherrick, C. E., & Cholewiak, R.W. (1986). Cutaneous sensitivity. In K. R. Boff, L. Kaufman,& J. L. Thomas (Eds.), *Handbook of Perception and Human Performance* (pp. 12-1 -- 12-58). New York, NY: John Wiley.

Sherry, R. R., & Ritter, F. E. (2002). *Dynamic Task Allocation: Issues for Implementing Adaptive Intelligent Automation* (Report No. ACS 2002-2). University Park, PA: The Pennsylvania State University School of Information Sciences and Technology.

Sinnett S., Spence C., & Soto-Faraco S. (2007). Visual dominanceand attention: The Colavita effect revisited. *Perception & Psychophysics*, 69(5), 673-686.

Smith, P. L., & Ratcliff, R. (2009). An integrated theory of attention and decision making in visual signal detection. *Psychological Review*, 116(2), 283-317.

Spain, R. D., & Bliss, J. P. (2008). The effect of sonification display pulse rate and reliability on operator trust and perceived workload during a simulated patient monitoring task. *Ergonomics*, 51(9), 1320-1337.

Spence, C. (2009). Crossmodal Attention. *Scholarpedia*. Retrieved March 13, 2010, from http://www.scholarpedia.org/article/Crossmodal_attention

Spence, C., & Driver, J. (1996). Audiovisual links in endogenous covert spatial attention. *Journal of Experimental Psychology: Human Perception and Performance,* 22(4), 1005-1030.

Spence, C., & Driver, J. (1997). Crossmodal links in attention between audition, vision, and touch: Implications for interface design. *International Journal of Cognitive Ergonomics*, 1(4), 351–373.

Spence, C., & Ho, C. (2008). Multisensory warning signals for event perception and safe driving. *Theoretical Issues in Ergonomics Science*, 9(6), 523-554.

Spence, C., & Santangelo, V. (2009). Capturing spatial attention with multisensory cues: A review. *Hearing Research*, 258(1-2), 134-142.

Stephan, K. L., Smith, S. E., Martin, R. L., Parker, S. P., & McAnally, K. (2006). Learning and Retention of Associations Between Auditory Icons and Denotative Referents: Implications for the Design of Auditory Warnings. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *48*(2), 288-299.

Stevens, S. S. (1968). Tactile vibration: Change of exponent with frequency. *Perception & Psychophysics*, 3, 223-228.

Summers, I. R., Cooper, P. G., Wright, P., Gratton, D. A., Milnes, P., & Brown, B. H. (1997). Information from time-varying vibrotactile stimuli. The *Journal of the Acoustical Society of America*, 102(6), 3686-3696.

Tadema, J., & Theunissen, E. (2008). Design of a synthetic vision overlay for UAV autoland monitoring. *Proceedings of SPIE*, 6957, 69570B-69570B-11.

Tripathi, P. (2008). Human-centric framework for perceptually adaptive interfaces. *Proceedings of the 22nd British CHI Group Annual Conference on HCI: People and Computers XXII: Culture, Creativity, Interaction*, 2, 255-256.

Van Erp, J. B. F. (2002). Guidelines for the use of vibrotactile displays in human computer interaction. *Proceedings of Eurohaptics 2002*, (pp. 18-22).

Van Erp, J. B. F. (2005a). Presenting directions with a vibrotactile torso display. *Ergonomics*, 48(3), 302-313.

Van Erp, J. B. F. (2005b). Vibrotactile spatial acuity on the torso: Effects of location and timing parameters. In *Proceedings of the First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems* (pp. 80-85).

Van Erp, J. B. F., & Van Veen, H. A. H. C. (2003). A multipurpose tactile vest for astronauts in the international space station. In *Proceedings of Eurohaptics 2003*, (pp. 405-408).

Van Erp, J. B. F., & Van Veen, H. A. H. C. (2004). Vibrotactile in-vehicle navigation system. *Transportation Research Part F: Traffic Psychology and Behaviour*, 7(4-5), 247-256.

Van Erp, J. B. F., Groen, E. L., Bos, J. E., & Van Veen, H. A. H. C. (2006). A tactile cockpit instrument supports the control of self-motion during spatial disorientation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 48(2), 219-228.

Van Erp, J. B. F., Van Veen, H. A. H. C., Jansen, C., & Dobbins, T. (2005). Waypoint navigation with a vibrotactile waist belt. *ACM Transactions on Applied Perception*, 2, 106-117.

Van Erp, J.B.F., & Self, B. P. (2008). *Tactile Displays for Orientation, Navigation and Communication in Air, Sea and Land Environments* (Report No. TR-HFM-122). Neuilly-sur-Seine Cedex, France: North Atlantic Treaty Organization, Research and Technology Organization.

Van Veen, H. A. H. C., & Van Erp, J. B. F. (2000). Tactile Information presentation in the cockpit. In *Proceedings of the First International Workshop on Haptic Human-Computer Interaction* (pp. 174-181).

Verrillo, R. T. (1962). Investigation of some parameters of the cutaneous threshold for vibration. *The Journal of the Acoustical Society of America*, 34(11), 1768-1773.

Verrillo, R. T. (1963). Effect of contactor area on the vibrotactile threshold. *The Journal of Acoustical Society of America*, 35(12), 1962-1966.

Verrillo, R. T. (1966). Effect of spatial parameters on the vibrotactile threshold. *The Journal of Experimental Psychology*, 71(4), 570-575.

Verrillo, R. T., Fraioli, A. J., & Smith, R. L. (1969). Sensation magnitude of vibrotactile stimuli. *Perception & Psychophysics*, 6(6A), 366-372.

Verrillo, R. T., Gescheider, G.A., Calman, B. G., & Van Doren, C. L. (1983). Vibrotactile masking: Effects of one- and two-site stimulation. *Perception and Psychophysics*, 33(4), 379-387.

Vicente, K. J. (2002). Ecological interface design: Progress and challenges. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 44(1), 62-78.

Vicente, K., & Rasmussen, J. (1992). Ecological interface design: Theoretical foundations. *IEEE Transactions on Systems Man and Cybernetics*, 22(4), 589-606.

Vroomen, J., Bertelson, P., & de Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics*, 63(4), 651-659.

Walker, B. (2002). Magnitude estimation of conceptual data dimensions for use in sonification. *Journal of Experimental Psychology: Applied*, 8(4), 211–221.

Walker, B., & Kramer, G. (2004). Ecological Psychoacoustics and Auditory Displays: Hearing, Grouping, and Meaning Making. In J. Neuhoff, *Ecological Psychoacoustics* (pp. 149-174). New York: Academic Press.

Walker, B. N., & Kramer, G. (2005). Mappings and metaphors in auditory displays. *ACM Transactions on Applied Perception*, *2*(4), 407-412.

Walker, B., & Kramer, G. (2006). Auditory Displays, Alarms, and Auditory Interfaces. *International Encyclopedia of Ergonomics and Human Factors*. Portland, Oregon, USA: CRC Press.

Watson, M., & Sanderson, P. (2007). Designing for attention with sound: Challenges and extensions to ecological interface design. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 49(2), 331-346.

Watson, M., Sanderson, P., & Anderson, J. (2000). Designing auditory displays for team environments. In *Proceedings of the 5th Australian Aviation Psychology Symposium (AAvPA)* (pp. 1-8).

Weinstein, S. (1968). Intensive and extensive aspects of tactile sensitivity as a function of body part, sex and laterality. In *Proceedings of the First International Symposium on the Skin Senses* (pp. 195-218).

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, 88(3), 638-667.

Wickens C. D. (1984). Processing resources in attention. In R. Parasuraman & D.R. Davies (Eds.), Varieties of attention (pp. 63-102). New York: Academic Press.

Wiener, E.L. (1989). *Human factors of advanced technology ("glass cockpit") transport aircraft* (NASA CR 177528). Moffet Field, CA: NASA Ames Research Center.

Williams, K. (2008). Documentation of Sensory Information in the Operation of Unmanned Aircraft Systems. Washington, DC, USA. Retrieved from http://www.faa.gov/library/reports/medical/oamtechreports/2000s/media/200823.pdf.

Wilska, A. (1954). On the vibrational sensitivity in different regions of the body surface. *Acta Physiologica Scandinavica*, 31(2-3), 285-289.

Wright, R., & Ward, L. (2008). *Orienting of Attention*. New York, NY: Oxford University Press.

Yanagida, Y., Kakita, M., Lindeman, R.W., Kume, Y., & Tetsutani, N. (2004). Vibrotactile letter reading using a low-resolution tactor array. In *Proceedings of the 12th International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems* (pp. 400-406).

Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: voluntary versus automatic allocation. *Journal of experimental psychology: Human perception and performance, 16*(1), 121-34.

Yantis, S., & Jonides, J. (1996). Attentional capture by abrupt onsets: New perceptual objects or visual masking?. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 1505-1513.

Annex A Annotated Bibliography

A.1 Ecological Interface Design

Reference: Burns, C. M. (2000). Putting it all together: Improving display integration in ecological displays. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 42(2), 226-241.
Overview: Burns explores how different methods of display integration affect the operator's understanding of an ecological display. She describes two different methods of grouping interface components with a display: spatial proximity and temporal proximity. Spatial proximity refers to groupings based on the spatial location of the item, while temporal proximity refers to grouping items by having them presented at the same time. Based on these two methods of organization items on an interface Burns hypothesized that "high-spatial and high temporal proximity of means-end related information will improve operator performance on fault diagnosis when compared with displays that do not keep means-end related information together in high-spatial/high-temporal proximity." Thus, operators would be able to integrate data from a display that is capable of indicating both high spatial and temporal proximity more effectively. The means-ends links between the data were generated using an EID approach. Three interfaces for a power plant were generated based on different combinations of spatial and temporal proximity (high-spatial & high-temporal (HH), high-spatial & low-temporal (HL), low-spatial & high-temporal (LH). The fourth possible combination (low-spatial & low-temporal) was not used since it was not feasible to design such an interface. Detection times, the time required until the fault was noticed, and diagnosis times, the time required until the problem behind the fault was identified, were recorded. The results of the experiment found that the HH display resulted in the fastest fault diagnosis times, and produced the largest amount of correct diagnoses (based on a 4-point ordinal scale used by Pawlak and Vicente (1996) which accounts for partial or vague diagnoses). However, the HL display did result in the fastest fault detection times (the amount of time until the operator first noticed that something was wrong with the plant, but before they've uncovered what has gone wrong). Burns concludes that spatial proximity helped improve operator performance, but temporal proximity was only relevant if spatial proximity was already present. The inclusion of temporal proximity did improve the response time of operators in diagnosis task, which Burns categorizes as a problem-solving task. Therefore, to provide maximum support to an operator on a difficult task that requires adaptive problem solving, integration should be supported using both temporal and spatial integration.
Conclusions: Integration of information in a multimodal interface is also an important concept. It is highly probable that different sources of information will be presented to different modalities (even if this information is only there to help orient attention), and the operator will need to assimilate

information from these different sources into a coherent picture of the system they are monitoring. The concepts of spatial and temporal proximity discussed in this paper are applicable when referring to a 2D visual display, however they may require revision if they are to be applied to a multimodal display. One of the reasons spatial proximity is important is because attention tends to be deployed to a specific set of stimuli, this can either be a spatial location, stimuli feature, or object (see Wright and Ward, 2008 for a recent review). In vision, there are also costs associated with scanning since saccades take time to re-orient the eyes. Thus, if two interface components can be inspected with overt orientation of attention, then perception would be more efficient. In a multimodal interface, most spatial cues are integrated into a cohesive attentional space centered on the individual, but there are special characteristics of how space is encoded that are modality specific which can interfere with attempts at creating spatial proximity.

Also, a multimodal interface may be designed with different spatial reference frames for information in different modalities. This is similar to how there are different frames of reference for the symbology and overlays on a heads-up display, and the outside environment. There is some evidence that there are conflicts in processing the two sources of information when different spatial frames of reference are available (Wickens and Hollands, 2000), and whether this conflict also occurs in a multimodal interface is still an open research question.

The methodology used to test the different interfaces within this experiment can be adapted for the design of UAV GCS experiments. In both cases, the operator is asked to identify when a fault has occurred with the system. While the participants in the GCS experiment will not be asked to explicitly diagnose the cause of the fault, they must respond with the correct abort response which requires some hypothesis of the cause of the off-nominal behaviour. In this experiment, analysis was done in terms of time until detection, time until diagnosis, detection accuracy, and diagnosis accuracy, which are variables that can be used in other experiments.

Reference:

Davies, T. C., Burns, C. M., & Pinder, S. D. (2007). Testing a novel auditory interface display to enable visually impaired travelers to use sonar mobility devices effectively. In *Proceedings of the 51st Annual Meeting of the Human Factors and Ergonomics Society* (pp. 278-282). Santa Monica, CA: Human Factors and Ergonomics Society.

Overview:

In this paper, Davies et al. discuss the design and preliminary testing of an auditory interface for a sonar mobility device for the visually impaired. The design of the interface was created using the EID framework in an attempt to provide the user with a larger set of information that they can use to assist with navigation. The goal of a sonar mobility device is to provide information about the spatial relationships between objects in the environment and the user, and the majority of these devices use very simple auditory interfaces which are based on *earcons* (“abstract musical tones that are represented in hierarchical forms to relay information). Unfortunately, these interfaces are limited because of their simplicity. One system (the Sonic Pathfinder) was only able to provide distance information for the nearest obstacle, and other relevant objects in the environment were not presented to the user.

Davies et al. conducted a WDA of the navigation task, and used the result analysis to create semantic mappings based on auditory icons and earcons. Auditory icons are sounds that represent an object or process that draws heavily from its real-world equivalent (Sanderson and Watson, 2005). In particular, they decided to use “Nomic” mappings for auditory icons, which link a sound to an external event in an intuitive manner. For example, they used the sound of footsteps as a way of representing moving obstacles, and they could change the amplitude or tempo of a sound as a way of representing size or speed. Normal earcons (pure tones which coded information) were also used as a comparison to the auditory icons. Three scenarios were used to test the ability of a participant to walk through a scenario and generate a report of what was in the environment. Sighted participants were used since previous studies had shown that there was not a large difference in the ability to perform localization exercises. The results of the experiment showed that participants benefitted from both types of auditory displays, but they were better at using the auditory icons (footsteps) than the earcons (abstract tones). A number of issues related to auditory localization such as front-back confusion were found, and it is important to note that individual head related transfer functions were not used.

Conclusions:

This study shows the effective use of an auditory modality to represent information that is normally communication through visual displays. In particular, spatial information about the environment was communicated using earcons and auditory icons. While this is possible, there are many problems with relying only on auditory information to communicate spatial information (e.g. the back/front and size comparisons can only be made when there is auditory stimuli available for objects which are being compared). Thus, when designing an interface where the designers have the option to switch between different modalities, care must be taken that the best modality is chosen for the task. While an EID approach was used in the design of the interface, very little discussion was done on how semantic information was mapped onto auditory properties other than a brief discussion of the differences between auditory icons and earcons. However, the finding that auditory icons do provide better performance than earcons does suggest that making use of intuitive mappings is a key part of designing a successful multimodal interface. Also, since this display was designed for a single modality, in the absence of cues from other modalities, many of the multimodal problems discussed in other papers (such as Watson & Sanderson, 2007) did not need to be addressed.

Reference:

Lee, J., Stoner, H., & Marshall, D. (2004). Enhancing interaction with the driving ecology through haptic interfaces. *IEEE International Conference on Systems, Man and Cybernetics*, 1, 841-846.

Overview:

This paper is the only published account of using the EID methodology for designing a haptic/tactile interface. Lee et al. discuss the motivation for using EID for analyzing and designing new in-vehicle information systems (IVIS). The most significant reason for using sound interface design methodologies for IVIS is that drivers would continue to use these systems even if they are distracting to the driving task. Lee et al. hope that EID could be used to ensure that the "driving ecology" is improved such that the information benefits of IVIS systems does not interfere with the task of driving safely.

However, the authors point at several differences between the driving environment and the process control scenarios that EID interfaces are normally used for. The most relevant (to UAV GCS and multimodal interfaces) differences are:

- Time frame: Many of the changes in typical application of EID "evolve over several minutes or hours", however in the driving scenario, events can occur in a matter of seconds, and judgements and responses must be made by the driver in that short timeframe.
- Degree of cognitive control for unanticipated events: In typical process control scenarios, unexpected events require diagnosis of faults using knowledge-based processing, however in driving scenarios there is an emphasis on *skill-based behaviour*.
- Perception of relevant information: In the driving domain, drivers are directly able to perceive the environment to gather a large portion of information that is relevant to the task. In the process control domain, most of the information is collected through sensors and displayed through remote interfaces.

Similar to the work done in the auditory modality using EID, Lee et al. explored some of the unique properties of haptic/tactile displays that must be considered when employing the EID methodology. The following table taken from the paper summarizes the major implications for interface design based on the level of cognitive control:

Table A- 1: Summary of the major implications for interface design based on the level of cognitive control (Lee et al., 2007, p. 843)

Level of cognitive control	Implication for haptic interface
Skill-based behavior—sensory-motor patterns guided by time-space *signals*.	Haptic signals should have a direct analogical link to the motor response requirements—people should be able to act directly on the displayed information.
	Components of haptic signals should be isomorphic with the components of movements they guide.
	Haptic signals should have a direct analogical link to the visual signals available to the driver.
	Haptic signals should direct driver attention to relevant information in the driving scene.
Rule-based behavior—pre-defined responses triggered by familiar *signs*.	Haptic signs should show the state of the system relative to goal-relevant invariants of driving.
	Haptic signs should provide salient cues to select appropriate sensory-motor patterns and to select appropriate pre-planned responses.
	Haptic signs should be based on abstract process properties that uniquely define the underlying system state.
Knowledge-based behavior—analysis based on interpretation of *symbols*.	Haptic symbols should represent the functional structure of the system at multiple levels of abstraction.
	Haptic symbols should represent functional relationships as perceptually accessible analogical interface features.

It is interesting to note that Lee et al. do not discuss the importance of *attentional mapping*, which is a key portion of the EID extensions proposed by Sanderson, Anderson, and Watson (2000). The driving domain requires frequent attention switches and multi-tasking, so an analysis of where attention *should* be directed would benefit this line of research.

Conclusions:

Since this paper is the only published account of extending EID to the tactile modality, it provides some initial insights into the benefits and challenges of using tactile and haptic displays. Similarly to the EID extensions found in the auditory EID literature, the authors suggest that *rule-based behaviour* should be supported by salient cues during transition points. The authors also suggest that skill-based behaviour should be supported by cues that have “a direct analogical link to the signals from the driving environment.” Since drivers are heavily immersed in the driving task, and have access to a variety of environmental stimuli, this is a valid suggestion. However, for process-control and remote operation of robotic vehicles, the signals may not be intuitive to the operators since the operators are remotely separated from the vehicle.

Reference:

Sanderson, P., Anderson, J., & Watson, M. (2000). Extending ecological interface design to auditory displays. In *Proceedings of the 2000 Annual Conference of the Computer-Human Interaction Special Interest Group (CHISIG) of the Ergonomics Society of Australia (OzCHI2000)* (pp. 259-266).

Overview:

In this paper, Sanderson et al. (200) provide a very thorough discussion of possible methods of extending the EID methodology to include auditory design elements. The authors discuss scenarios where switching to the auditory modality may provide performance increases over the visual modality. These include low cognitive load vigilance tasks, where auditory displays have been proven to improve performance, high cognitive load tasks, where the auditory modality is able to provide additional information and draw on different attentional resources, tasks where vision is overloaded, and tasks where it is disadvantageous to shift attention away from a location.

While it is very tempting to include auditory information into each of these contexts, improper use of auditory displays can also be distracting and lead to decreased performance. Therefore, Sanderson et al. turn to the EID methodology to ensure that the additional information presented using the auditory modality is relevant and useful to the operators. When considering EID framework, the authors propose three questions that must be addressed:

1. How can the principles of EID clarify when to present information visually or auditory?
2. Is EID an adequate theoretical framework for guiding the design of auditory displays, or does it need to be extended?
3. Do we have the necessary knowledge about auditory processes to guide the design of auditory displays?

Some of these questions are partially addressed within the paper. To help guide the modality of information presentation, Sanderson et al. propose the use of attention mapping as well as semantic mapping. They also suggest that the EID methodology should draw from other parts of cognitive work analysis (CWA) to assist with the attention mapping step. The following table describes how different steps in CWA can be applied to auditory interface design.

Table A- 2: Applying CWA phases to auditory interface design (Sanderson et al., 2000, p. 262)

CWA phase	**Description**	**Issues for auditory displays**
Work Domain Analysis (WDA) • Functional purpose • Priorities and values • General function • Physical function • Physical form	Provides information about why the system or work domain exists, the flow of information or value through it, its functions, and the physical processes and objects underlying its functions.	Helps to identify work domain characteristics and relations that need to be displayed in any interface. For example, physical properties of work domain may indicate candidates for *audification*. Information is necessary but insufficient for interface design at this point.
Control Task Analysis • Temporal coordination control task analysis (TC-CTA) • Control task analysis (CTA)	Provides information about what needs to be done in the work domain, by whom, when, and how information about activity might be transmitted. Also gives information about temporal relations between tasks	In helping to identify a temporal profile of ongoing tasks, and possible competition between tasks, CTA leads analysts to knowledge about an appropriate *attentional profile* across tasks. This leads to conjectures about which tasks are best displayed visually, and which auditorily.
Strategy analysis (SA)	Provides information about different ways, if more than one way exists, in which the control tasks can be carried out.	Range of strategies available to human controllers may be extended by considering the possibilities of auditory displays in an interface.
Social organisational analysis (SOA)	Provides information about how work is shared across multiple actors in a complex organisation and how multiple actors coordinate efforts	Indicates where auditory display might help or hinder coordination between actors, given the *obligatory* nature of most auditory displays.
Worker competencies analysis (WCA)	Provides information about the form of cognitive control needed for a task, distinguishing skill- rule- and knowledge-based behavior.	Indicates intrinsic or training-based characteristics of workers that might point to the effectiveness of auditory elements in interface displays. Auditory display and especially *sonification* may help move cognitive control towards SBB.
Semantic mapping (SM)	Provides information about criteria for choosing interface elements so that goal-relevant task invariants are mapped onto key perceptual properties of the interface's behavior.	Gives designers a framework for judging the information-carrying potential of dimensions of an auditory stimulus, based in a knowledge of *auditory perception*.
Attentional mapping (AM)	Provides information about whether and when a control task should be supported in focal or non-focal attention.	Gives designers requirements for how an auditory display should control attention alongside other interface elements, based in a knowledge of *auditory attention*.

Sanderson et al. also explore various possibilities of mapping data to auditory characteristics in the semantic mapping step of EID. To assist with this process they turned to seven guidelines for presenting visual information. They discussed auditory equivalents for some of these heuristics:

1. Goal achievement as figural goodness. Sanderson et al. equated the concept of figural goodness in visual stimuli to acoustic simplicity.
2. Work domain constraints as visual containers. Containers are a spatial concept that Sanderson et al. state is difficult to replicate in the auditory domain.
3. Process dynamics as figural changes. In the auditory domain this could be represented by changes in acoustic parameters.
4. Functional relations as visual connections. Relationship of different acoustic parameters to each other.

The other three heuristics were: pictorial symbols to represent components, alphanumerical output where needed, and time as visual perspective.

Finally, Sanderson et al. also discuss the importance of understanding how auditory attention is shifted between different auditory stimuli. They felt that an auditory display would need to be usable both when attention is directed to it, as well as when it is in peripheral attention. Since auditory displays are obligatory, special care must be taken to manage the operator's attention.

Conclusions:

This paper provides a wealth of information about auditory displays, and how EID can be extended to assist in the development of multimodal interfaces. The three questions proposed by the authors are very relevant to extending EID to beyond visual and auditory displays. The first question addresses whether EID is capable of providing guidance about which modality information should be presented. Currently, EID does not provide any guidance on this matter beyond the suggestions presented in this paper. The WDA is typically done without any consideration about the form of the final interface; its goal is to provide a list of variables and relationships between the variables. Thus, the designer must decide how to present these variables through the interface. One possible method that EID can be extended to help address this problem is by using a data classification system. In the paper, the authors state that the auditory modality is especially suited for displaying temporal information. Some variables may have characteristics that are more dynamic (varies with time) than static, and these variables may benefit from being presented through an auditory signal. This represented how a characteristic of the variable provides guidance of the modality of its presentation, and this additional step of data classification can be built into the EID methodology. The second question is answered by the authors through their suggestions of adding an attentional mapping stage to EID, as well as drawing from other analyses done in CWA. The use of attentional mapping and task information becomes very relevant when temporal aspects of data presentation are considered. Finally, the last question posed by Sanderson et al. can also be applied to the tactile modality, but the details about how this can be done still needs to be investigated.

Reference:

Sanderson, P. M., & Watson, M. O. (2005). From information content to auditory display with *ecological interface design*: Prospects and challenges. In *Proceedings of the 49th Meeting of the Human Factors and Ergonomics Society* (pp. 259-263). Santa Monica, CA: Human Factors and Ergonomics Society.

Overview:

This paper continues the research done previously by Sanderson, Anderson, and Watson (see Watson, Sanderson, and Anderson, 2000; Sanderson, Anderson, and Watson, 2000) and explores how EID can be extended to auditory interface design. Specially, they examine the use of a *visual thesaurus* (Burns and Hajdukiewicz, 2004) and discuss the prospects and challenges of creating an analogous auditory thesaurus. The visual thesaurus is a set of visual forms that can be used to represent work domain properties. The visual forms used include visual primitives (bar graphs and other simple iconic elements) and, complex combinations of visual primitives (connections, grouping, etc.). By using these individual elements a "visual ecology" can be created which allows the operator to process information about system constraints based on visual perceptual judgements.

Sanderson and Watson extend this process to auditory displays by first considering what types of visual primitives are available to a designer: *auditory icons*, *earcons*, *audifications*, and *sonifications*. *Auditory icons* are sounds that represent a fact or situation that draws heavily from its real-world equivalent. An *earcon* represents a “discrete sound that is a member of a set of sounds that are related to each other through a syntactic structure”. *Audifications* are a translation of some physical stimuli into an auditory representation. Finally, a *sonification* is the mapping of a source, or multiple sources in the world into auditory dimensions of an auditory signal. These sound primitives can serve as building blocks for an auditory display. For example *skill-based behaviour* must be supported using a display that consists of a space/time signal. Thus, Sanderson and Watson state that *audifications* and *sonifications* would be able to facilitate this level of cognitive control.

Finally, Sanderson and Watson discuss a number of challenges that still need to be addressed when using EID to support the design of auditory displays because auditory displays are ubiquitous, obligatory, and transitory, while visual displays are localized, optional, and persistent. These challenges can be separated out into two types of problems (the paper describes it as 4 challenges): those that relate to the distribution of information amongst team members (who needs the information), and those that relate to the temporal distribution of information (when should information be displayed).

<u>Conclusions:</u>

There are a number of points made by the authors which are applicable to the design of multimodal interfaces. The use of a visual thesaurus has helped streamline the process of using EID for visual displays. Therefore, auditory and tactile thesauruses could help multimodal interface designers in a similar manner. The authors propose a design process at the end of the paper which can be followed by those who are designing a multimodal interface. It is important to note that these design guidelines also consider how information needs to be distributed across a team of individuals. This is an element not usually considered in EID, which tends to have a focus on single user displays.

1. Who needs to keep track of which part of the work domain?
2. What is the sensory context of the work domain (visual, auditory, haptic)?
3. What variables and relations need to be displayed?
4. When and how fast do variables change and how should this be mapped to displays?
5. What level of cognitive control is needed?
6. Which modality or modalities would provide the most natural mapping for the task?
7. Is there an existing design pattern that would fit the above requirements?
8. For visual displays:
    a. Provide framework for visual displays.
    b. Provide details in a process like that of Burns and Hajdukiewicz (2004).
    c. Test the results.
9. For auditory displays:
    a. Perform attentional mapping across different people in the workspace.
    b. Perform attentional mapping for the primary person who will monitor

c. Test the results.
10. Test the combined effect.

There are three elements that could be improved in this design process which were not included by Sanderson and Watson. First, each modality is considered separate, and interaction effects are only considered at the end. For a fully multimodal interface, some information could be displayed in informative elements that span modalities (though there is no research in this area as of yet). Secondly, while EID normally considers the complete redesign of an interface, there may be portions of the work domain that cannot be changed (a piece of equipment or a standard that must be used in the redesign), these static pieces may constrain the other design elements and should be included as part of the design process. Finally, some of the steps have unclear goals. For example, the “best” modality may be interpreted as fastest response time, highest accuracy, or easiest to learn, depending on the task supported.

Reference:

Sarter, N. B. (2006). Multimodal information presentation: Design guidance and research challenges. *International Journal of Industrial Ergonomics*, 36(5), 439-445.

Overview:

In this paper Sarter discusses a number of design guidelines for multimodal interfaces, as well as challenges that still need to be addressed for multimodal interfaces to become mainstream. Multimodal interfaces have become increasingly used in systems because multimodal presentation can provide synergy, redundancy, and increased bandwidth of information transfer. To assist with this task, a number of design guidelines have been produced. However, Sarter states that these guidelines vary in their applicability and focus: some focus on information presentation in a single modality and focus on sensory characteristics of that modality, while others are higher-level guidelines that work across multiple modalities. These guidelines often do not take advantage of the underlying perceptual and neurophysiological research that has been done, or they are re-iterations of existing guidelines that are not well justified or explained. Sarter describes a number of challenges that are not adequately addressed by the existing guidelines:

- *Modality expectations*: If an operator expects a cue to appear in a certain modality, they experience “enhanced readiness to detect and discriminate information in that sensory channel.” This may lead to situations where individuals are “tunnelled” into one modality, leading to missed targets in non-expected modalities.
- *Modality shifting effect*: Operators have difficulty shifting their attention away from an expected modality to a modality that contains less frequent targets.
- *Crossmodal attention shifting*: Shifts in spatial attention in one modality also tend to shift attention in other modalities.
- *Exogenous and endogenous attention*: In real-world tasks, an operator will have goal-driven (endogenous) responses to stimuli, but the interface is also able to capture attention using stimuli-driven (exogenous) cues. The interaction between these two forms of attention is still not well understood.

Sarter also summarizes existing guidelines into four topics:

- *Selection of modalities:*
  - Determine if multiple modalities are necessary.
  - Investigating benefits of using multiple modalities compared to the costs of increased interface management.
  - Environmental constraints (ambient noise, vibrations) must be noted and considered.
- *Mapping of modalities to tasks and types of information:*
  - Characteristics of the data should map onto characteristics of the modality used.
  - Some characteristics of each modality are presented within the paper.
- *The combination, synchrononization and integration of modalities:*
  - There are few guidelines that deal with the "resulting spatial and temporal combination and synchronization of these channels" and many of these are conflicting. Some suggest minimizing the overlap between modalities, while others state that it should be based on user preferences.
  - User preferences for multimodal combinations can be detrimental in team environments and because it puts added responsibility on the operator.
  - System and context functionality should also guide how multimodal data is presented.
- *The adaptation of multimodal information presentation:*
  - It is not possible to have fixed assignments of modalities to specific tasks or types of attention.
  - An adaptive interface can reduce some of the additional interface management costs of multimodal interfaces.

<u>Conclusions:</u>

This paper is an excellent review of the current state of multimodal interface research and has a focus on application. Any extensions to EID should take care to address the challenges that were presented in this paper. It is also important to present design guidelines that are rooted in the scientific literature to help justify the design decisions. One thing to note is that Sarter states that many of the crossmodal effects are relatively small (small differences in reaction time), and the author hypothesizes that these differences could become larger in a real-world application. However, the opposite may also be true: the effect sizes may be small enough that they disappear in a real-world system because they are overcome by other factors. The research on endogenous and *exogenous attention* may provide some further insight into this question.

<u>Reference:</u>

Vicente, K. J. (2002). Ecological interface design: Progress and challenges. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 44(1), 62-78.

<u>Overview:</u>

This paper reviews a number of findings related to the use of EID, and describes the benefits of the framework as well as some challenges that still need to be addressed (as of 2002). Vicente describes that the role of a human in a complex sociotechnical system is to be a "knowledge worker by engaging in adaptive problem solving." He also explains that an EID design is built using a *work domain analysis* instead of a task analysis because the unexpected events that interfaces are designed to support are not sufficiently captured in a task analysis. This is accomplished using the abstraction hierarchy (AH) and skills, rules, knowledge (SRK) taxonomy. A review of the literature examined how the EID framework impacts operator performance. The results showed the EID systems provided performance increases in terms of increased speed at resolving faults, and decreased variability in results. These findings were found to be greatest for complex situations that required adaptive problem solving, which support the claims made by the EID framework. These claims are that EID supports the skill based, rule based, and knowledge based behaviour of the user. The literature reviewed by Vicente did not show any improvements in performance for complex tasks. However, he states that this is not a drawback because operators were able to achieve similar levels of performance "despite the added visual complexity compared with traditional designs." Vicente also examined why the performance advantages for EID existed. He found that EID interfaces provided benefit by restructuring the information using the AH which allowed operators to focus on the functional goals of the system. This allowed "higher level control", because operators could monitor the system at a very high level without delving into the details. Vicente mentioned that these benefits existed outside of the new visual forms used in EID interfaces, but the new visual forms did improve performance by loading spatial resources instead of verbal resources. As a last point, Vicente mentioned that there existed large individual differences in the ability for an operator to make use of the EID system. However, in a study by Sharp and Helmicki (1998) less experienced users (residents) in an experiment in the medical domain received greater benefit than the more experienced users (attending physicians). The participants in that study were asked to make diagnoses using displays that contained either functional information and graphics forms, or traditional alphanumeric data. A number of challenges were also proposed by Vicente. One such concern, which has been more recently addressed in Burns and Hajdukiewicz (2002), is that EID provides little guidance for the actual implementation of the interface components. This challenge could partially be addressed by using interface design principles that are complimentary to the EID framework to provide a systematic method for the design of the interface. A secondary issue is the lack of guidance for thedesign of a display's layout. Visual momentum (techniques that help reduce the disorientation a user might feel as the interface transitions through different screens) and spreading the display of information across multiple monitors is difficult because the entirety of the AH should be visible to the operator. Vicente also acknowledged that the EID framework has largely been focused on visual displays, but the fundamental elements that make the technique effective are not restricted to this modality.
<u>Conclusions:</u> Vicente provides a broad overview of the benefits of the EID framework, and some limitations related to its implementation. The benefits cited are largely related to reorganizing the types of information that are displayed to the operators and would be applicable to displays in any sensory modality. It is only the spatial versus verbal loading benefit that may differ across modalities.

Both the auditory and tactile modalities allow for the orientation of spatial attention to certain locations, but they also allow for processing advantages that are not spatial. For example, an operator could make use of perceptual judgements based on changes in pitch of an auditory signal that do not make use of its spatial characteristics. Also, information in the tactile modality can be communicated through changes in amplitude, frequency, duration, or through patterns. Thus, the processing advantages should be thought of in terms of perceptual judgements, which may include non-spatial discriminations, versus analytical judgements which rely more heavily on short-term memory.

The concepts of visual momentum and the distribution of information across multiple displays are also very applicable to the design of multimodal interfaces. As additional modalities are included as part of the interface, the designers are given a larger “display space” through which to communicate to the operator. Instead of having information distributed through multiple spatial locations in the visual modality (as is the case with multiple displays), the information is now distributed through multiple spatial and perceptual locations in multiple modalities. The research presented about crossmodal attention will be vital in understanding how best to support the transition between the different channels of information that the operator will be presented with.

Reference:

Vicente, K., & Rasmussen, J. (1992). Ecological interface design: Theoretical foundations. *IEEE Transactions on Systems Man and Cybernetics*, 22(4), 589-606.

Overview:

This paper outlines the theoretical foundations of the ecological interface design (EID) framework. Vicente and Rasmussen state that the complex systems that operators are often required to control require a special type of interface because of three reasons: complex systems require complex controllers, physical systems are governed by constraints, and finally that good controllers must possess a model of a system. They propose the EID framework as a method for designing such an interface.

This framework is built on top of Rasmussen’s skills, rules, and knowledge taxonomy which separates tasks and goals into different levels of cognitive control. *Skill-based behaviour* (SBB) represents behaviour that exists in situations that are common during operation. Due to extensive training and experience, operators are able to respond almost automatically using learned motor-skills. *Rule-based behaviour* (RBB) occurs in situations that do not occur as often but are foreseen by the designers of the system. In these cases, the designers are able to design rules and procedures that the operators should follow during off-nominal events. Finally, *knowledge-based behaviour* (KBB) exists when events that are unforeseen by both of the operator and designers occur. Since there are no set procedures that can be referred to, the operators must rely on their knowledge of the system to improvise a solution.

The EID framework attempts to support the operator in skill, rule, and knowledge based behaviour by making use of three principles:

"1) SBB- To support interaction via time-space signals, the operator should be able to act directly on the display and, the structure of the displayed information should be isomorphic to the part-whole structure of movements.
2) RBB- Provide a consistent one-to-one mapping between the work domain constraints and the cues or signs provided by the interface.
3) KBB- Represent the work domain in the form of an abstraction hierarchy to serve as an externalized mental model that will support knowledge-based problem solving." Taken from Vicente & Rasmussen 1992

Vicente and Rasmussen provide evidence to support adoption of these principles as the foundations of EID. Firstly, operators tend to resort to lower levels of cognitive control (SBB and RBB), even when the interface does not naturally support control at this level. This is because the lower levels of cognitive control are less effortful than KBB. However, operators also make use of KBB when support for the lower levels of cognitive control is included in an interface. Thus, all three levels of cognitive must be available to operators so that operators can use the simplest level of cognitive control that is needed for a task.

Secondly, decisions made by perceptual judgements have less variability than ones made using analytical judgements. The perceptual judgements take advantage of perceptual processes that are highly efficient and specially tuned to detect certain changes. Therefore, if an interface was designed to show system constraints through the use of perceptual constraints, the operator could make use of RBB. Vicente and Rasmussen state that an operator using this kind of control would exhibit KBB while making use of RBB.

Finally, an abstraction hierarchy (AH) is a method of providing multiple levels of understanding a system that are joined together using a goal-oriented (means-ends) method. By building a representation of the AH into the interface, the entire scope of the problem space is visible to the operator. This external representation of the entire system and its boundaries, assist the operator with KBB. The levels of abstraction in the AH also allow the operator to move between different levels of understanding the system, helping them control the complexity of the system. This allows them to keep track of higher level goals when things are normal, while providing the ability to drill down to the details when things go wrong and KBB is required.

Conclusions:

The most direct connection to multimodal interfaces is through principle 2. Vicente and Rasmussen argue that interfaces where "domain invariants are mapped isomorphically onto perceptual invariants" reduces the variability of responses, and allows the operator to rely on RBB. In this paper, the focus was on the design of visual display, thus many of the perceptual invariants discussed relate to configural displays and Gestalt psychology. These are well established visual ideas, however when designing for other modalities, analogous perceptual invariants in the auditory and tactile domain must be discovered. In the visual domain, the focus is on building a larger perceptual form out of smaller elements, and this is larger due to the nature of the AH. It is still unknown if this is the correct approach to take when approaching the other modalities. However, the use of perceptual judgements as a way of monitoring a rule is still a valid method of reducing the amount of work that the operator needs to do.

The first principle also has implications for the design of multimodal interfaces because many of the examples of skill based behaviour make use of feedback in non-visual modalities. Many motor-tasks rely solely on proprioception (Wickens and Hollands, 2000), and musicians and singers are often trained to respond to *auditory signals* (matching a pitch and harmonizing are two examples). Switching to other sensor modalities gives the designer access to different types of signals that can carry information, and some of these may have advantages over visual signals that are commonly used.

<u>Reference:</u>

Watson, M., & Sanderson, P. (2007). Designing for attention with sound: Challenges and extensions to ecological interface design. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 49(2), 331-346.

<u>Overview:</u>

This paper elaborates on the design process first discussed in Sanderson and Watson (2005) and applies it to the anaesthesia monitoring scenario discussed in Watson, Sanderson, and Anderson (2000). In this scenario, an anaesthesiologist must monitor a patient's vital signs in an operating room. Many of the cues which they must monitor for are visual, and some are based on observation of the patient's body. Thus, the visual modality was heavily overloaded with the addition of visual monitors. The design process discussed previously was refined for use in a real-world design application. A graphical representation of this process is shown below.

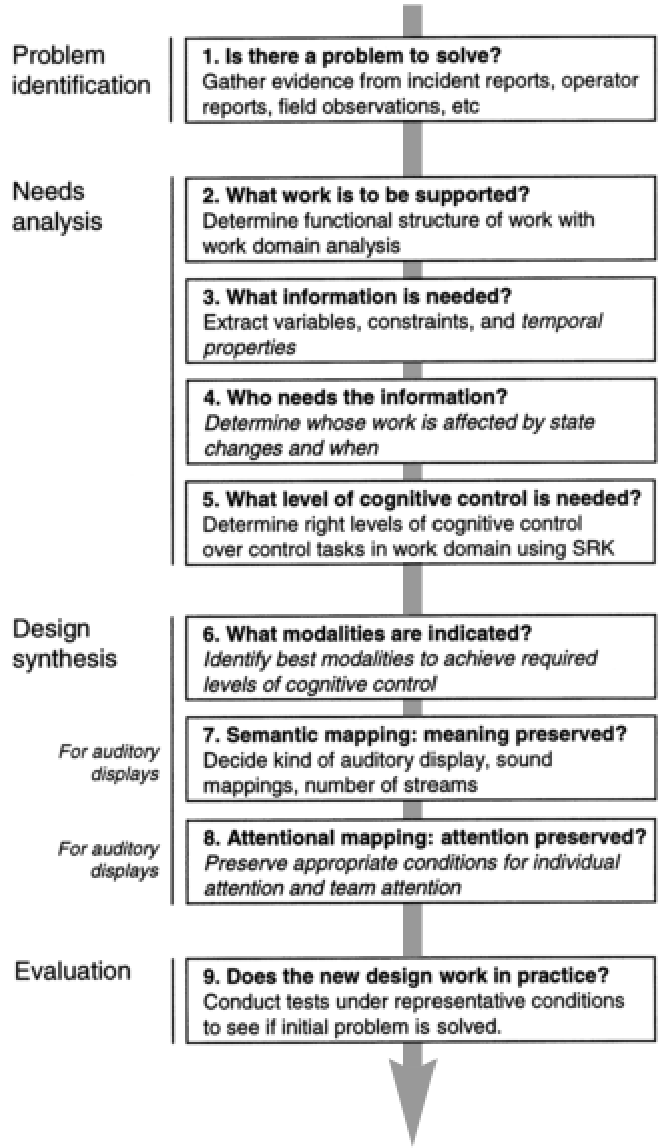


Figure A- 1 Auditory EID design process (Watson and Sanderson, 2007, p. 2)

The major differences between this design process and the original proposed design process is the inclusion of the problem identification step and a formal evaluation step. However, less emphasis is put on testing for possible expected or unexpected crossmodal effects. This design process was followed in the design of an anaesthesia monitoring device for use in operating rooms. There were a couple of specific design problems that applied to this design domain. First, many of the higher order variables in the abstraction hierarchy could not be directly sensed or measured by sensors. Since they could not be sensed or measured, there is a gap in information required for the abstraction hierarchy and the operator is not able to understand the entire abstraction of the system However, Vicente and Rasmussen (1992) proposed that the higher order variables could be calculated either using equations or models, but Watson and Sanderson suggest that due to the unreliability of the sensors, it would be misleading to display only calculated higher order variables. Instead, lower order data is shown which the operators can integrate into higher order variables, while also monitoring the sensors' reliability. Another problem was determining the level of abstraction to display. Visual displays can often show information at multiple levels by using a large number of separate displays which the operator can navigate through. This was not possible with the auditory display being designed, so a single "view" had to be chosen, and this required intimate knowledge of the domain, and an understanding of the tasks involved.

The final design of the sonification involved a number of design decisions. During the semantic mapping step, these design decisions involved determining the kind of auditory display, candidate sound dimension mappings, and the number of auditory streams. In the attention mapping stage, both individual and team attention was considered. For individuals who are actively monitoring the sonification, the goal was to make the sonification capture the listener's attention when the data was crossing a boundary condition. For the team, care was taken to ensure that the sonification was not overly distracting for team members who were not actively monitoring the stream. The initial evaluations of the auditory interface created using EID were positive. The authors found that the sonification supported the skill-based behaviour that they envisioned. These results suggest that the suggested extensions to EID assist with the systematic design of an auditory interface.

Conclusions:

The design process described in this paper can be adapted for use with other modalities. However, it is still unclear how this can be accomplished. This is due to the large amount of information that is specialized in each modality. Therefore, further work must be done to highlight similarities that exist between the different modalities in an attempt to see if the EID design process can be generalized to multimodal design. If this is not possible then special design processes for each modality may be required.

Reference:

Watson, M., Sanderson, P., & Anderson, J. (2000). Designing auditory displays for team environments. In *Proceedings of the 5th Australian Aviation Psychology Symposium (AAvPA)* (pp. 20–24).

Overview:

In this paper, Watson et al. (2000) explore methods for designing auditory displays for team environments. The authors state that auditory information can play a large role in information presentation to operators, but most traditional interfaces have focused on using visual outputs. When auditory outputs are being used, they are normally implemented through devices such as alarms which are adjunct to visual displays. However, auditory alarms have possible drawbacks because the auditory modality cannot be "shut out" and its perceptual processing is obligatory. Auditory alarms which are annoying, possibly due to high false alarm rates, or distracting because of their sound levels, are often physically turned off because they cannot be "tuned out". Instead, the authors state that *sonifications* should be used as a method for reducing the intrusiveness of data presented aurally. The *sonification* can be used to ensure smooth transitions between focal awareness and peripheral awareness of the stimuli, and would provide useful information in both states, as seen in the figure below.

Exploiting Auditory Attention (modified from Sanderson et. al., 2000)

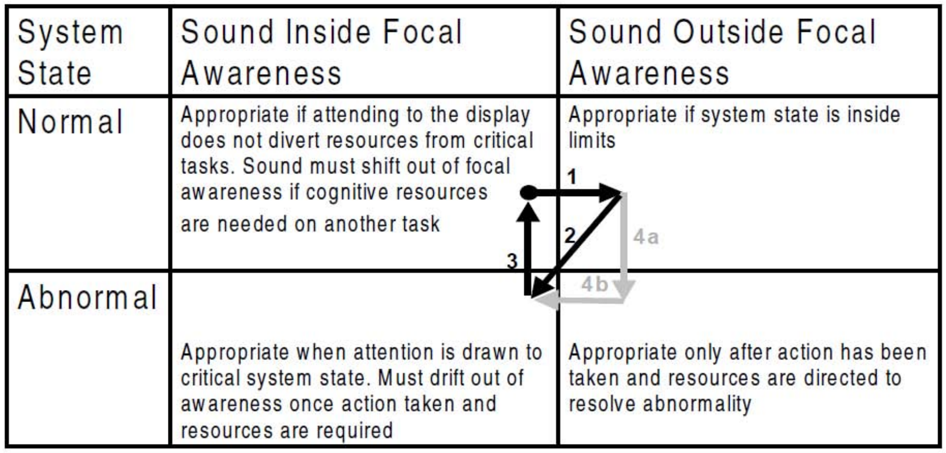


Figure A- 2: Exploiting Auditory Attention (modified from Sanderson et al., 2000, p. 265)

Watson et al. use the EID framework as a way of determining what information should be presented using the auditory modality. In addition to the normal work domain analysis (WDA) that outlines the constraints on a system (Vicente & Rasmussen, 1992), Watson et al. also describe the usefulness of other portions of Cognitive Work Analysis (CWA) that are particularly useful for team-displays. They also propose an additional attentional mapping stage, largely based on the analysis done in the later stages of the CWA (control task analysis, strategy analysis, and social organization analysis) to be important for multimodal and team-based displays. This is because operators may not always be focused on the display, thus it is crucial to know where their attention is directed. The following table highlights these points.

Table A- 3: Issues for CWA when designing auditory displays (Watson et al., 2000)

Issues for CWA when designing auditory displays.

CWA phase	**Issues for auditory displays**
Work domain analysis	Identifies domain characteristics and relationships to be displayed in any interface.
Control task analysis	Identifies a profile of ongoing tasks, competition between tasks and *attentional profiles* across tasks. Auditory or visual display?
Strategy analysis	Auditory displays may extend the range of strategies available.
Social organizational analysis	Indicates where auditory displays might help or hinder coordination between actors, given *obligatory* nature of most auditory displays.
Worker competence analysis	Indicates characteristics of workers that might point to the effectiveness of auditory elements in interface displays.
EID steps	
Semantic mapping	A framework for judging the information-carrying ability of dimensions of an auditory stimulus based on knowledge of *auditory perception*.
Attentional mapping	How an auditory display should control attention alongside other interface elements, based on a knowledge of *auditory attention*.

Two possible applications were also discussed by the authors: auditory displays in the operation room and auditory displays to assist with approach and landing in an airplane cockpit. Of the two, the landing scenario is more relevant to our project. Watson et al. conducted a brief CWA on the landing scenario, and used this analysis to complete semantic and attentional mappings for the information required. The Work Domain Analysis (WDA): an analysis of the objects, relationships, and constraints within a work domain) and CTA revealed two major types of variables: those related to spatial location (altitude, air speed and direction), and those related to "engineering function" (control of thrust and automation). These variables were mapped onto different variables of the sonification, with the spatial variables mapped onto the spatial origin of the sonification, and engineering function variables mapped onto properties of the sonification (such as mapping speed to the tempo of the sound, and direction of thrust as a harmonic interval). The attentional mapping elements of the analysis were not described.

Conclusions:

This paper provides two important elements for the design of multimodal interfaces, even though the focus was on presenting information through the auditory modality. The first is the use of both semantic and attentional mapping as part of the EID framework. When working with a single modality, such as vision, a designer can largely assume that the operator's attention will be focused on the display. However, as the number of channels of information increases, the assumption of focused attention is no longer valid. This is true even for purely visual displays that are spread out over many monitors, or if a task also requires observation of non-display elements in the environment. The control-task analysis and strategy analysis is important because it provides some guidance on what tasks are occurring, and priorities that the operator may have.

The second important element is the figure describing the transitions of focal and peripheral attention relative to variable normality. It can also be applied to tactile displays because the tactile modality is also an information channel that cannot be "shut out". As a consequence of this, it is important to discover which elements of the auditory and tactile modality (and to a lesser extent the visual modality) can be perceived pre-attentively and which require focal attention to process.

A.2 Tactile Perception

Reference:

Brewster, S. A., & Brown, L. M. (2004). Tactons: Structured tactile messages for non-visual information display. In *Proceedings of the 5th Australasian User Interface Conference* (pp. 15-23). Sydney, Australia: Australian Computer Society.

Overview:

This paper introduces *tactons*, also known as tactile icons, as brief tactile messages that can be used to represent complex concepts and information in a vibrotactile display. They are tactile counterparts of icons. The general basic parameters (such asfrequency, amplitude, waveform, rhythm, etc.) that can be controlled to create tactons are discussed in this paper. Three types of tactons are introduced:

1. Compound tactons: A combination of two or more simple tactons. A simple tacton can be a vibration which has been generated by a single parameter, like high or low frequency vibration.

2. Hierarchical tactons: A node in a tacton tree which inherits properties from the tacton (node) located in at higher level above it.

3. Transformational tactons: Present several properties by encoding each property by means of a tactile parameter. For example, if a transformational tacton is used in a mobile phone, the type of the alert (voice call or text message) can be encoded by rhythm and the priority of the alert can be encoded by amplitude.

Conclusions:

Tactons can be considered as one of the options to present complex information in a vibrotactile display.

Reference:

Brown, L. M., Brewster, S. A., & Purchase, H. C. (2005). A first investigation into the effectiveness of tactons. In *Proceedings of the First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems* (pp. 167-176). Washington, DC: IEEE Computer Society.

Overview:

Two experiments were performed to investigate the design of tactons. Two vibrotactor devices were used in this research: the Audiological Engineering Corporation (AEC) TACTAID VBW32 transducer and the Engineering Acoustics Inc (EAI) C2 Tactor.

Methodology:

The first experiment was run to investigate whether subjects could differentiate between different amplitude modulated signals in terms of roughness. Five stimuli were used in this experiment. For generating the stimuli a 250Hz sine wave was chosen as the base signal and it was modulated by 20, 30, 40 and 50 Hz signals.

The first experiment was run twice, once for each of the vibrotactors. Amplitude modulated signals were presented to the participant’s index finger through vibrotactors. The experimental task of the subjects was to compare two stimuli and indicate which stimulus felt “rougher”. Every possible pairing of stimuli was presented four times.

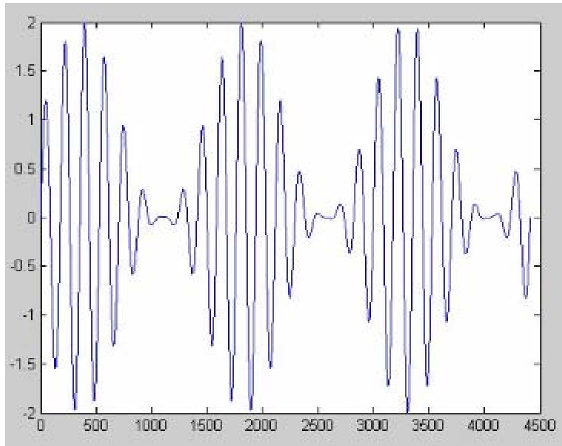


Figure A-3: A 250 Hz sinusoid signal modulated by a 30Hz sinusoid signal. Figure taken from Brown et al. (2005, p.3).

The second experiment investigated the effectiveness of tactons in conveying abstract messages. Vibrations with different durations can be grouped together to create rhythmic units. For this part experiment, three types of alerts (voice call, text message and multimedia message) were encoded using different rhythms. The priority of these alerts (low, medium or high) was encoded using different roughness levels. As an example, the same rhythm was used to present a high priority text message and low priority text message, but they were presented with different roughness levels.

Results:

The results of the first experiment "indicated that participants felt that roughness increased as modulation frequency decreased, with the exception of the un-modulated sine wave, which felt less rough than all other stimuli". The results of the first experiment showed that the C2 Tactor was found to be more reliable in providing different levels of roughness. Therefore, the second portion of the experiments was done only with this vibrotactor.

As the result of the second experiment, the average discrimination rates of 93% and 80% were recorded for alert types (represented by different rhythms) and priority of alerts (represented by different roughness levels) respectively. The average result for overall tacton recognition was 71%.

Conclusions:

Subjects can differentiate between different amplitude modulated signals in terms of roughness. Feeling of roughness increases as the modulation frequency decreases. The results of the experiments demonstrated that the C2 tactor is a suitable vibrotactor for creating tactons. Tactons can effectively convey complex messages to operators in a very concise manner in vibrotactile displays.

Reference:

Cholewiak, R. W., Brill, J. C., & Schwab, A. (2004). Vibrotactile localization on the abdomen: Effects of place and space. *Perception & Psychophysics*, 66(6), 970-987.

Overview:

A series of experiments were executed to investigate the effects of how the placement of vibrotactile stimuli affects localization on the torso.

Methodology:

In the first part of the experiment, stimuli were presented using vibrotactors situated at 12 equidistant locations on two belts. The belts encircled the abdomen and the lower margin of the rib. The reason for using two levels (abdomen and lower margin of the rib) was to see "whether the characteristics of the underlying tissue would affect the localization of the vibrotactile stimuli or not?" The vibrotactors located on the frontal side of the lower belt fell over the tissue of the belly, whereas vibrotactors of the upper belt were over the ribs. In each trial, one stimulus (vibrotactor) was activated.

In the second part of the experiment, the number of vibrotactors on the belt decreased to eight and six in order to reach better possible localization performance.
In the third part of the experiment, 7 vibrotactors were located on a short strip spanning roughly half the circumference of the body and this tactor strip was used in 4 locations on the torso: front, back, left side and right side of the body. In the first case the array across the abdomen (front) was arranged so tactor 1 was at the left, tactor 4 at the navel and tactor 7 at the right side. For the back case, tactor 1 was at the right side, tactor 4 at the spine and tactor 7 at the left side of the body. The other two cases had similar orientations, but had tactors that started at the navel or spine, and a center tactor (4) on either the left or right side of the body.

Results:

The results of the first portion of the experiment revealed that the performance of detecting stimuli around the abdomen and the rib cage was similar. Therefore for the torso, the underlying tissue type plays a minor role in vibrotactile spatial location. The ability to localize a stimulus around the torso was found to be a function of proximity to the spine (6 o'clock) and the navel (12 o'clock). It was found that observers were more capable of correctly detecting stimulus near the spine (6 o'clock) and the navel (12 o'clock) and these points can serve as anatomical reference points for the trunk.

Results of the second part of the experiment showed that performance was dramatically improved when the number of vibrotactors was reduced.

As the results of the third part of the experiment, better performance was obtained when the tactor strip was used on the front and back, rather than when it was located on the left or right side of the body.

Conclusions:

The underlying tissue type plays a minor role in vibrotactile spatial localization on the skin of the torso. The spine and the navel can work as natural anchor points. Observers are more capable of correctly detecting and localizing stimulus near these points. Increasing tactor separation in a vibrotactile display will improve the localization performance of the users. In case of using a vibrotactor strip spanning half the circumference of the body, better performance can be obtained when the tactors span the front or the back side of the body when compared to when tactor strip span the left or right side of the body.

Reference:

Cholewiak, R. W., & Collins, A. A. (2000). The generation of vibrotactile patterns on a linear array: Influences of body site, time, and presentation mode. *Perception & Psychophysics*, 62(6), 1220-1235.

Overview:

Influences of timing parameters and presentation modes on the generation of vibrotactile patterns were investigated in a set of experiments.

Methodology and results:

In this study, patterns were presented to the distal pad of the left index finger, the left forearm and the lower back region by means of seven vibrotactors for each area. Two modes of pattern presentation were used; saltatory and veridical. In the veridical mode, all of the seven vibrotactors that were situated in a linear array were activated in sequence to provide a linear pattern. In the saltatory mode, seven bursts of vibration were presented at only three tactor sites. Three bursts of vibration were presented through the first; three bursts through the fourth; and one burst through the seventh vibrotactor. The vibrations were presented in the two modes with different *Burst Durations* (*BD*) and Inter Burst Intervals (*IBI*). The values for the *BDs* and the *IBIs* were 4, 9, 17, 26, 35, 70, and 139 msec.

Two experiments were run. The main goal of the first experiment was to find out “how efficiently can a good line be generated?” For this part, subjects were instructed to rate the levels of perceived length, smoothness, spatial distribution and straightness of the patterns.

The results of the first experiment showed that when vibrations were presented with longer *Bs*, subjects perceived longer lines. Significant interaction between *BD* and *IBI* was also found. With longer *IBIs* for stimuli with a given *BD*, the generated lines were felt to be longer. This means that as velocity of activation sequence increases, the perceived length of patterns decreases. The stimuli were perceived to be smoother with shorter *IBIs.* Perceived smoothness of patterns was found to be mainly a function of *IBI*. Perceived Spatial distribution was reported to have better quality when small *BDs* and *IBIs* were used. Finally judgments of straightness improved with shorter *BDs* and shorter *IBIs.* This means that the velocity increment of an activation sequence will result in judgments of straighter patterns.

Because of similar judgments of subjects over the different body parts in the first experiment, for the second experiment, vibrations were presented only to the lower back.

The aim of the second experiment was to find out “to what extent subjects can discriminate the difference between two presentation modes (veridical and saltatory), and which of these modes can specifically generate a better line?” The second experiment was run in two parts.

For the first part, participants were instructed to judge “whether the line produced by a pair of stimuli were perceived to be same or different?” For this part, in half of the trials the modes of presentation were same and in half they were different.

For the second part of the second experiment, pairs of stimuli were presented to the participants and the presentation mode was always different for the two stimuli. Subjects were instructed to judge “which of a pair of stimuli generated a better line?”

As the result of the first part, when same stimuli were presented, 82% of the responses were correct. When stimuli were presented in two different modes, only 37% of the responses were correct.

The results of the second part of the experiment revealed that the veridical mode was superior to the salutatory mode, but the differences were very small.

Conclusion

A linear pattern can be generated by sequentially activating vibrotactors which are situated in a linear array. Linear patterns can be used to intuitively present information regarding orientation or direction in a vibrotactile display. When using a row of vibrotactors to represent messages that include a vibrotactor line , we should remember that:

1. As the velocity of an activation sequence increases, the perceived length of the line decreases.
2. The perceived smoothness of the line can be improve with shorter IBIs
3. Increase in velocity of activation sequence will result in judgments of straighter lines.

Reference:

Craig, J. C. (1972). Difference threshold for intensity of tactile stimuli. *Perception & Psychophysics*, 11(2), 150-152.

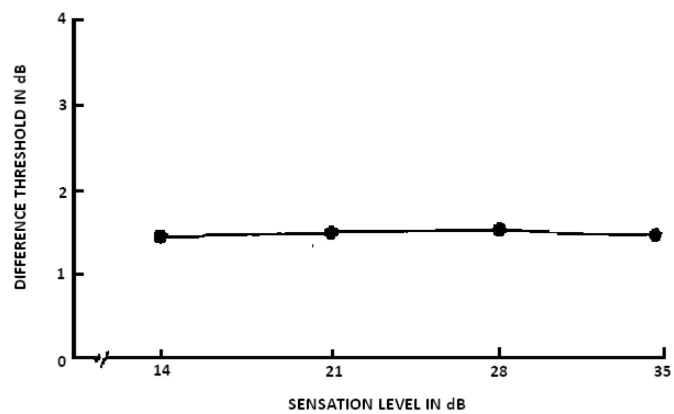
Overview:

Difference threshold (discriminated change in amplitude) for different intensity levels of tactile stimuli were measured in the presence and absence of a background noise. A 160 Hz vibration with 200ms duration was presented to the right index finger in the presence and absence of background vibration (noise).

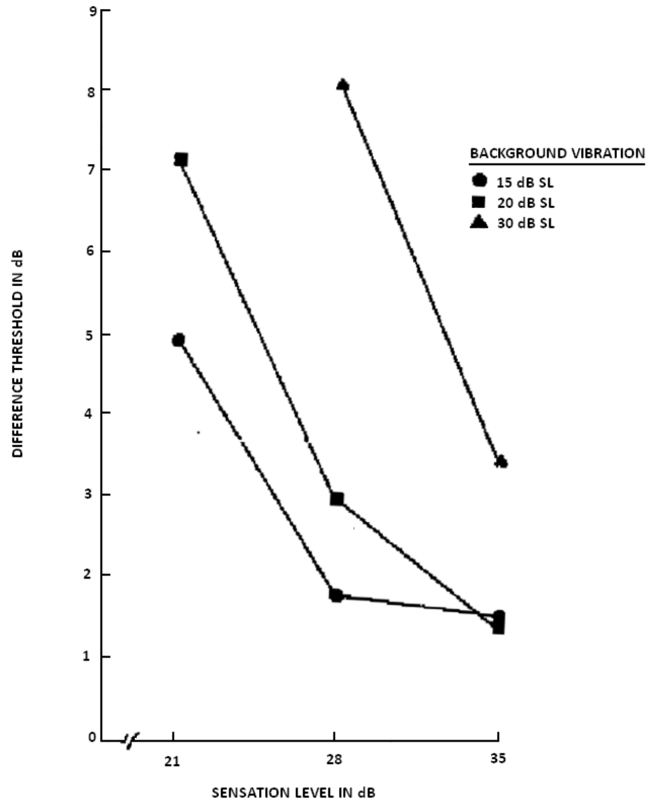
Methodology and results:

For the case in which the background vibration was absent, a 160 Hz vibratory signal was presented to the right index finger of the subjects. The difference threshold was measured for the vibration intensity levels of 14, 21, 28 and 35 dBSL. It was found that the difference threshold at these levels is constant and is approximately 1.5 dB in absence of background vibration.

When the background vibration was present, the subjects were presented with two 500 ms bursts of vibrations (as background vibration) and the 160 Hz signal were presented at the center of each burst. In this case, the difference threshold was measured for the vibration intensity levels of 15, 20 and 30 dBSL. It was found that the background vibration increases the difference threshold of the vibratory signal. Results of the experiment are illustrated in *Figure A-4*.



(a)



(b)

Figure A-4: Difference threshold for vibrotactile stimuli in absence (a) and presence (b) of a background vibration. Figures taken from Craig (1972, p.150-151).

Difference thresholds increased as the intensity of background vibrations was increased.

<u>Conclusions:</u>

When designing patterns in a vibrotactile display we should remember that background vibrations increase the difference threshold of the vibrotactile stimuli intensity. They can have negative influence on effectiveness of vibrotactile patterns.

Reference:

Craig, J.C., Evans, P.M.(2000). Vibrotactile masking and the persistence of tactual features. Perception & Psychophysics, 42(4), p 309-317

Overview:

Two experiments were executed in order to investigate the persistence of effects of forward maskers in *temporal masking*. Temporal masking occurs when the vibrations are presented to the same location, and the target stimulus is presented either within the time interval of the masking stimulus, or near the onset or just after the offset of the masking stimulus. *Forward masking* occurs when the target stimulus is corrupted with a preceding masking stimulus.

Methodology:

In the first experiment a masker pattern was presented to the subjects followed by a target pattern and subjects were instructed to ignore the first pattern and recognize number of lines in the second (target) pattern. Patterns were presented to the left index fingerpad of the subjects. The vibrotactile display consisted of 144 vibratory pins arranged in a 24 × 6 array (1.1 cm in width and 2.7 cm in height). Patterns were constructed by vertical or horizontal lines of vibration. Each line was made up of two rows or columns of pins. For example letter "F" consisted of three lines. The masker was generated by activating all of the 144 pins simultaneously. Figure A-5 illustrates the representations of the patterns which were used in this experiment.

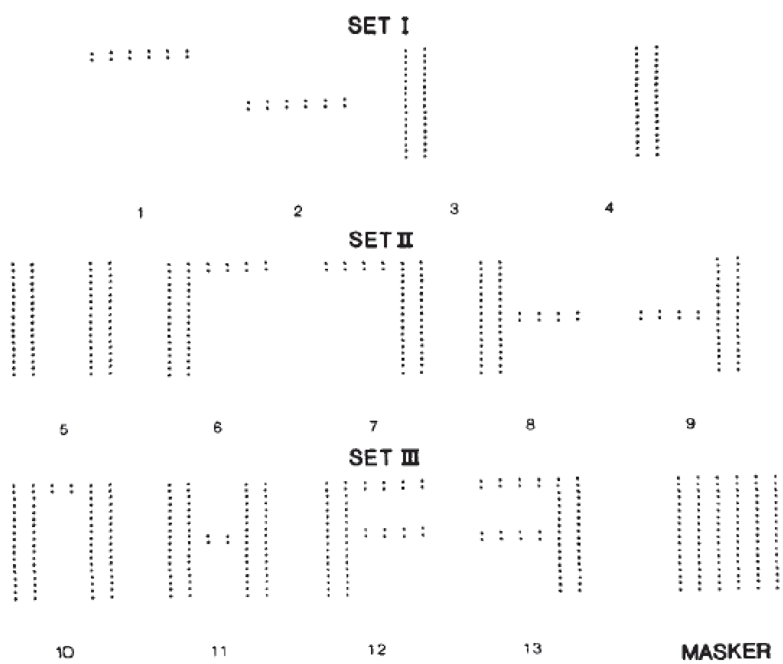


Figure A-5: Representations of the patterns and the masker. Figure taken from Craig and Evans (1987, p.310).

Results:

The results of the experiments revealed that at briefer *SOAs* there was more *backward masking* than forward masking. As *SOAs* increased, forward masking decreased more gradually than backward masking. At long *SOAs* there was more forward than backward masking. Forward masking remained visible for *SOAs* up to 1200 ms. Figure A-6 shows the results of the experiments and compares them to the results of the another study(Evans and Craig 1986) which was done to investigate the persistence of effects of backward maskers.

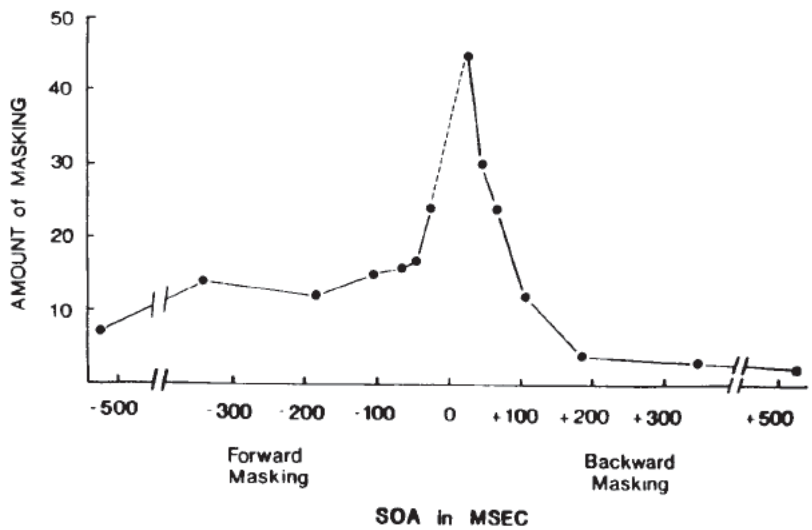


Figure A-6: Amount of forward and backward masking as a function SOA. *Values are percent correct in identification of a target pattern in the absence of a masker minus the percent correct in identification of a target pattern in the presence of a masker. Figure taken from Craig and Evans (1987, p. 311).*

It is obvious from Figure A-6 that Forward and backward masking have their greatest amounts at *SOAs* below 100 ms.

Conclusions:

Masking effects may have negative influence on perception of tactile patterns. Therefore, we should be aware of masking properties when designing vibrotactile patterns:

1. Forward and backward masking have their greatest levels at *SOAs* below 100 ms.
2. As *SOAs* increases, forward masking decreases more gradually than backward masking.
3. At briefer *SOAs* there is more backward masking than forward masking

Reference:

Jones, L. A., Lockyer, B., & Piateski, E. (2006). Tactile display and vibrotactile pattern recognition on the torso. *Advanced Robotics*, 20(12), 1359-1374.

Overview:

This paper describes two experiments regarding vibrotactile pattern recognition on the trunk. In the first experiment, the ability of subjects to identify eight different vibrotactile patterns was investigated. Patterns were presented to the lower back of the subjects by means of a 4×4 tactor array. The subjects were navigated through a path which had been designated by a grid of cones.

Methodology and results:

A 4×4 tactor array was mounted on the lower back of the participants. The distance between the rows of tactors was 4 cm and the column spacing was 6 cm. During the first experiment, subjects were seated on a stool. They were trained with the vibrotactile patterns before the experiment initiation. Each pattern was presented 3 times during the training. Figure A-7 illustrates the patterns which were used in the experiment. The results indicated that subjects were capable of discriminating all of the patterns with almost perfect accuracy.

In the second part of the experiment, the ability of subjects to recognize the same patterns while they were used as navigation commands were examined. In this portion of the study subjects were navigated through a path using the same tactor patterns of the first part of the experiment. For example, pattern D from Figure A-7 was used to represent the command “turn left”. The path had been designated by a grid of cones.

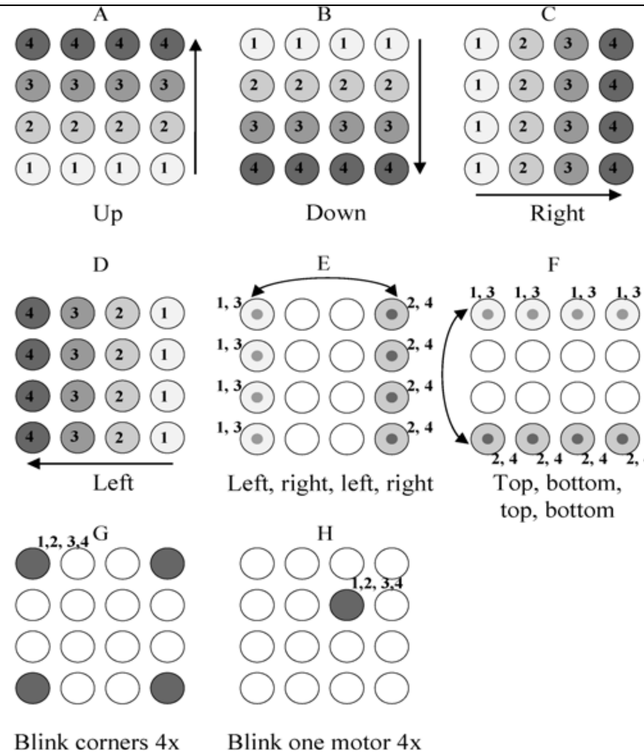


Figure A-7: The vibrotactile patterns generated by means of a 4×4 array of vibrotactors. Arrows represent the spatial order of activation. Figure taken from Jones et al. (2005, p. 1367).

Subjects were able to accurately follow the navigation commands and walk through the course using only the vibrotactile patterns as navigational commands.

Conclusions:

The results of the experiments demonstrated that vibrotactile spatio-temporal patterns presented to the torso can be recognized with high accuracy. Therefore, these patterns can be considered as reliable option to provide navigational information to operators through vibrotactile display.

Reference:

Kaaresoja, T., & Linjama, J. (2005). Perception of short tactile pulses generated by a vibration motor in a mobile phone. In *Proceedings of the First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems* (pp. 471-472). Los Alamitos, CA: IEEE Computer Society.

Overview:

This study investigated the user perception of vibrations generated by a mobile phone device.

Methodology and result:

Six different lengths (12.5, 25, 50, 100, 200, 500 ms) of mobile phone vibrations were presented to the group of subjects in three different locations: hand, trouser front pocket and belt.

Figure A-8 illustrates the results of the experiment for the case in which the mobile phone were located in the front pocket of the subjects. The results for the other locations were similar to this case. However, the pulses with 12.5 and 25 ms durations were slightly better perceived in hands. When the pulses lengths were 100 ms they were not judged as very strong vibrations, whereas when the pulses lengths were 500 ms, about 35% of the times they were judged to be strong and irritating.

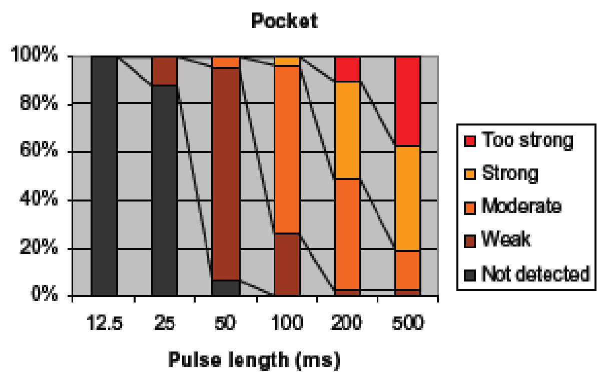


Figure A-8: Scoring of subjects for the perception of mobile phone vibrations located in their front pocket. Figure taken from Kaaresoja and Linjama (2005, p.2).

The results indicated that the duration of vibratory alerts should be between 50 and 200 ms. Vibrations shorter than 50 ms may not be sensed and vibrations longer than 200ms were reported to be irritating.

Conclusions:

Duration of vibratory alerts should be between 50 and 200 ms. Vibrations shorter than 50 ms may not be sensed and vibrations longer than 200ms were reported to be irritating.

Reference:

Kirman, J. (1974). Tactile apparent movement: The effects of interstimulus onset interval and stimulus duration. *Perception &Psychophysics*, 15(1), 1-6.

Overview:

Effects of *Stimulus Onset Asynchrony* (*SOA*) and stimulus duration on spatio-temporal integration (vibrotactile apparent movement) were investigated in this study. Judgements of apparent movement can be generated by sequentially activating a series of vibrotactors which are situated in an array (Cheung, Van Erp, and Cholewiak, 2008).

Methodology:

For this experiment the vibratory stimuli were presented to two different locations on the right index finger. The vibrations were varied in both duration and the inter-stimulus onset interval (*SOA*). They were presented in 6 durations (1, 10, 20, 50, 100, and 200ms) and were combined with each of 10 *SOAs* (10, 20, 30, 50, 70, 90, 110, 130, 150, and 200ms). Therefore a total of 60 pairs of stimuli were presented to the subjects. Participants were instructed to judge and rate the quality of the perceived apparent movement.

Results:

As the result of this experiment, it was found that the quality of perceived apparent movement varies as a function of *SOA*. Figure A-9 shows this function for stimuli with duration of 200ms. Considering this figure, the two stimuli provide the best feeling of apparent movement when the inter-stimulus onset interval was approximately equal to 130 ms. This means that the second stimulus started to stimulate after 130ms from onset of the first stimulus. In fact the two stimuli had a 70ms overlap.

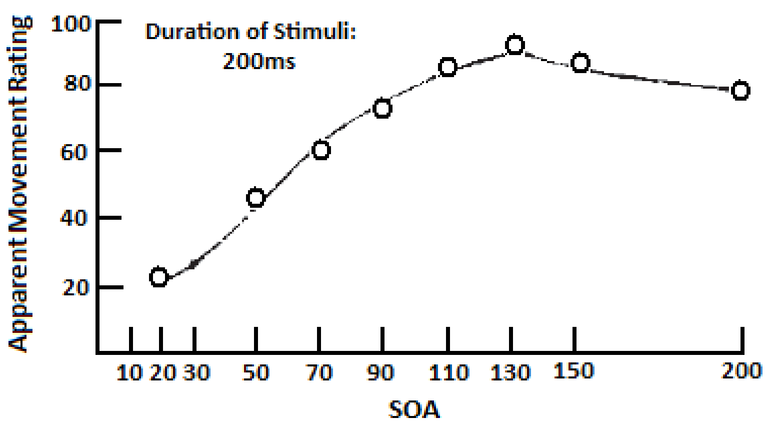


Figure A-9: Apparent movement rating as function of SOA (Results of Kirman experiment). Figure taken from Kirman (1974, p.2).

Figure A- 10 shows the optimal *SOAs* for different stimuli durations applied in the experiment. According to this figure, the optimal *SOA* for stimuli with durations of 1, 10, 20, 50, 100 and 200 ms to be perceived as an apparent movement are approximately 70, 50, 50, 70, 90 and 130 ms respectively.

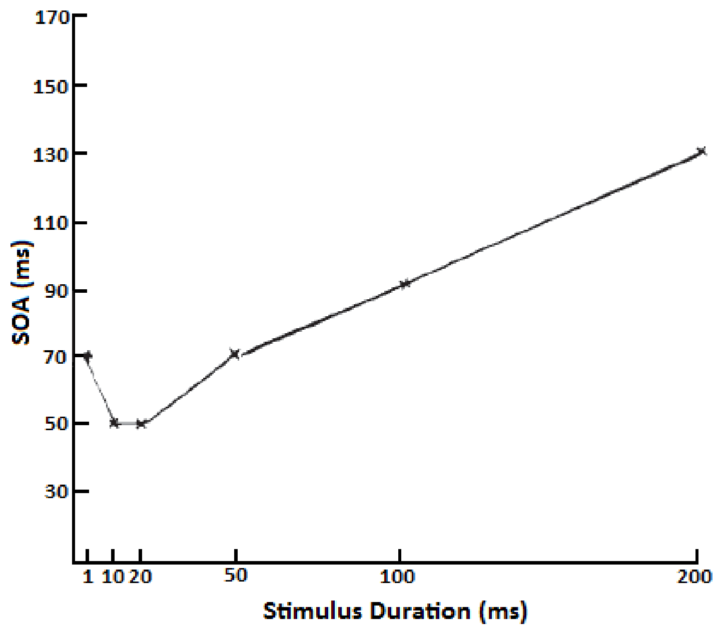


Figure A- 10: Optimal SOA as a function of stimulus duration (Results of Kirman experiment). Figure taken from Kirman (1974, p.3).

Finally, Figure A-11 shows the judgments of apparent movement for the optimal *SOAs* as a function of stimulus durations. According to this figure, as stimuli duration increases, judgments of apparent movement increase for optimal *SOAs.* As the result of this study, we can conclude that when spatio-temporal patterns are being used in vibrotactile displays, the quality of perceived apparent movement is a function of inter-stimulus onset interval (*SOA*) and *burst duration* (*BD*).

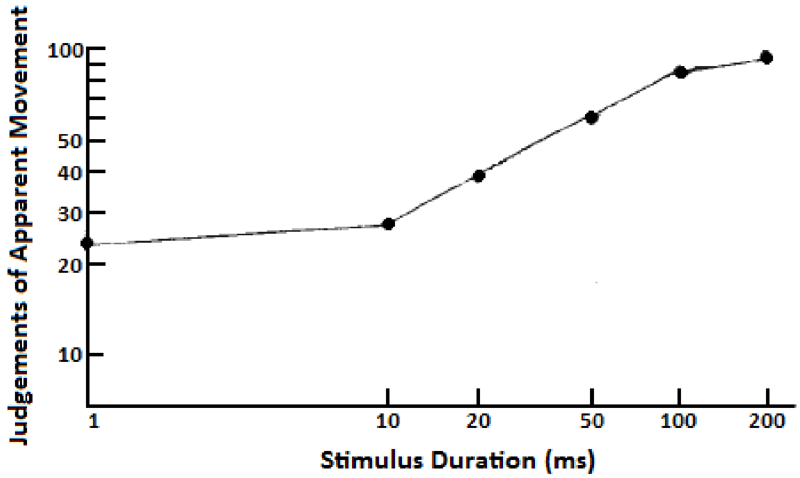


Figure A-11: Judgments of apparent movement for the optimal SOAs as a function of stimulus duration (Results of Kirman experiment). Figure taken from Kirman (1974, p.5).

Conclusions:

When spatio-temporal patterns are being used in vibrotactile displays as a way to convey information, the quality of perceived apparent movement is a function of inter-stimulus onset interval (*SOA*) and *burst duration* (*BD*).

Reference:

Lindeman, R. W., & Yanagida, Y. (2003). Empirical studies for effective near-field haptics in virtual environments. In *Proceedings of the 2003 IEEE Virtual Reality Conference* (pp. 287-288). Los Alamitos, CA: IEEE Computer Society.

Overview:

In this experiment, the ability of subjects to localize a vibrotactile stimulus in a 3×3 tactor array was investigated.

Methodology:

The vibrotactor array was affixed to the backrest of an office chair, such that vibrations were presented to the lower back region of the subject's torso. The spacing between centers of each pair of neighbouring tractors was 6 cm. The tactors at the lowest row touched the back of the subjects just above the belt line. The centre column of the array was arranged along the subject's spine. 36 trials were executed for each participant. The experimental task of the subjects was to localize the vibration through a visual interface software which was designed for the experiment.

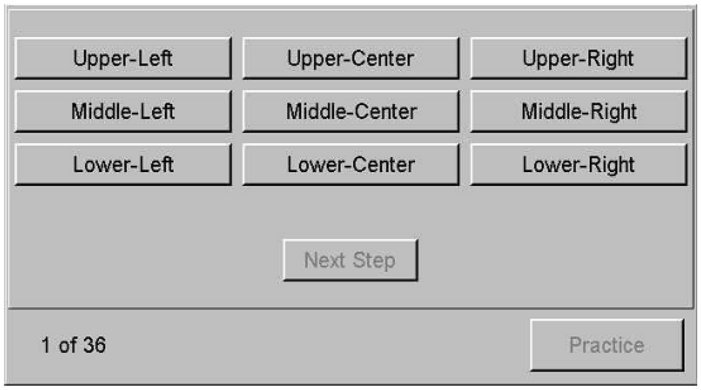


Figure A-12: The visual interface which was designed for the experiment. Figure taken from Lindeman and Yanagida (2003, p.2).

Results:

An overall accuracy of 84% correct localization rate was recorded as the result of this experiment. Vibrations presented to the upper row of the tactor array were more mislocalized than the other rows. There was no difference between the two lower rows.

Conclusions:

A vibratory stimulus presented to the back can be localized with relatively high accuracy and reliability.

Reference:

Rupert, A. H. (2000, March-April). An instrumentation solution for reducing spatial disorientation mishaps. *IEEE Engineering in Medicine and Biology Magazine*, 19(2), 71-80.

Overview:

Engineering solutions to deal with spatial disorientation mishaps in cockpits are presented in this paper. Vibrotactile displays were used in this study. They consisted of an array of vibrotactors which were embedded in a garment torso and could be worn by pilots. The garment was fabricated of stretchy textile material to maintain pressure between the tactors and the skin.

The Tactile Situation Awareness System (TSAS) was developed to control the tactors of the tactor locator system (TLS). A series of TLS prototypes were fabricated and worn by rotary-wing and fixed-wing pilots. A number of flight tests were executed to determine to what extent a pilot can intuitively maintain normal orientation and control when using the TSAS. The results of the flight tests are reported in detail in this paper.

General results demonstrated that TSAS prototypes were excellent tools to counter spatial disorientation.

Conclusions:

Tactile cues can assist pilots in spatial orientation during situations in which they can become disoriented.

Reference:
Sherrick, C. E. (1985). A scale for rate of tactual vibration. *The Journal of the Acoustical Society of America*, 78(1), 78-83.

Overview:

Two experiments were run in order to provide a scale for frequency of vibratory stimuli.

Methodology and results:

Vibratory stimuli were presented to the left index finger of subjects. Frequency of vibration was varied in ten steps from 2 to 290 Hz (2, 4, 6, 10, 20, 32, 54, 105, 183, and 290 Hz). The intensity of pulses was varied in three steps: 20, 28 and 36 dBSL. The experimental task of the participants was to assign a number corresponding to the perceived frequency of vibration.

The results for the estimates of frequency of vibrations as a function of actual frequency are plotted in Figure A-13.

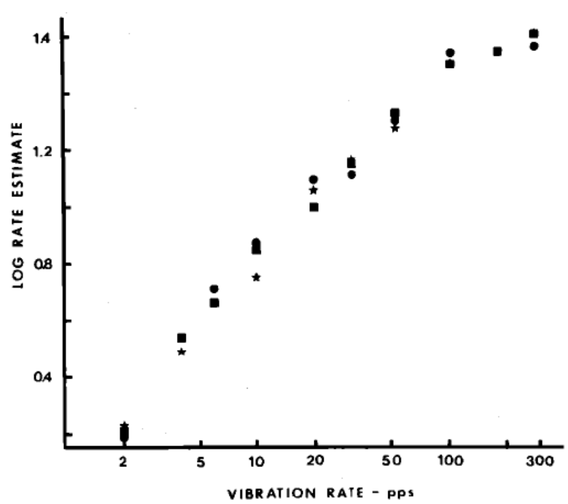


Figure A-13: Estimates of vibrations on the fingertip for 10 steps of vibration frequencies. Three levels of intensity were used for each step. Circles at 20dB, squares at 28 dB SL and stars at 36 dB SL. Figure taken from Sherrick (1984, p.80).

Considering Figure A-13, although vibrations were presented in three intensity levels, no significant effect of intensity is evident. This figure shows that discrimination of frequency steps plateau above 100Hz.

In the second portion of the experiment, stimuli were presented in the same frequency steps, but the intensity levels were different for each step. Table 1 shows the frequency of presented vibrations and their magnitudes. Subjects were asked to rate the level of the perceived vibration by pressing one of keys in a ten button touch tone pad (1 to 10).

The results of the second portion of the experiment revealed that a low frequency vibration at high intensity can be incorrectly perceived as a moderate vibration at medium intensity. This confirms the fact that increment of amplitude of a vibration increases the perceived frequency of the signal.

Table A-4: Frequency of vibrations and their magnitudes. Table contents are taken from figure 4 of the Sherrick (1984, p. 81).

Frequency of vibration (Hz)	Intensity level (Magnitude) (dB SL)
2	20
4	28
6	36
10	20
20	28
32	36
54	20
105	28
183	36
290	20

Conclusions:

Information can be encoded through vibrations with different frequencies or amplitudes in a vibrotactile display. For example, different levels of urgency can be presented by means of different levels of frequency or amplitude. When using amplitude or frequency parameters of vibration to present information in a vibrotactile display, we should always remember that:

1. A low frequency vibration at high intensity may be incorrectly perceived as a moderate vibration at medium intensity.
2. Increment of amplitude of a vibration increases the perceived frequency of the signal.
3. There is a correlation between frequency and amplitude of a vibratory stimulus.

Reference:

Stevens, S. S. (1968). Tactile vibration: Change of exponent with frequency. *Perception & Psychophysics*, 3, 223-228.

Overview:

The equal sensation functions for vibrations at different frequencies were investigated:

Methodology and results:

The equal sensation functions for vibrations at different frequencies were investigated by two methods:

1. Matching by adjustment: In this method, subjects were presented with two vibratory stimuli on their middle finger. They were instructed to adjust the level of the variable stimulus by means of a potentiometer such that its magnitude appeared equal to the reference stimulus.

   At the beginning of each session, each participant was asked to adjust the stimulus intensity at a just detectable level (for each of the three frequencies to be worked with). By this way, subjects determined the sensation threshold of stimuli.

   After the threshold determinations, one vibration at a specific frequency was set by the experimenter at one of three amplitude levels and the participants had to adjust the level of the variable stimulus to produce an apparent match. Figure A-14 shows all the matches that had a 60 Hz vibration in common. The 60 Hz vibration was used either as the reference stimulus which was adjusted by the experimenter (unfilled symbols), or as the variable stimulus adjusted by the subjects (filled symbols). The matches that had a 125 Hz vibration in common are shown in Figure A-15. It should be noted that some matches were repeated in two sessions: 60 Hz vibration matched to 30 Hz vibration, and 15 Hz vibration matched to 60 Hz vibration and 30 Hz Vibration matched to 125 Hz vibration. According to the results, the repeatability was reasonably good.

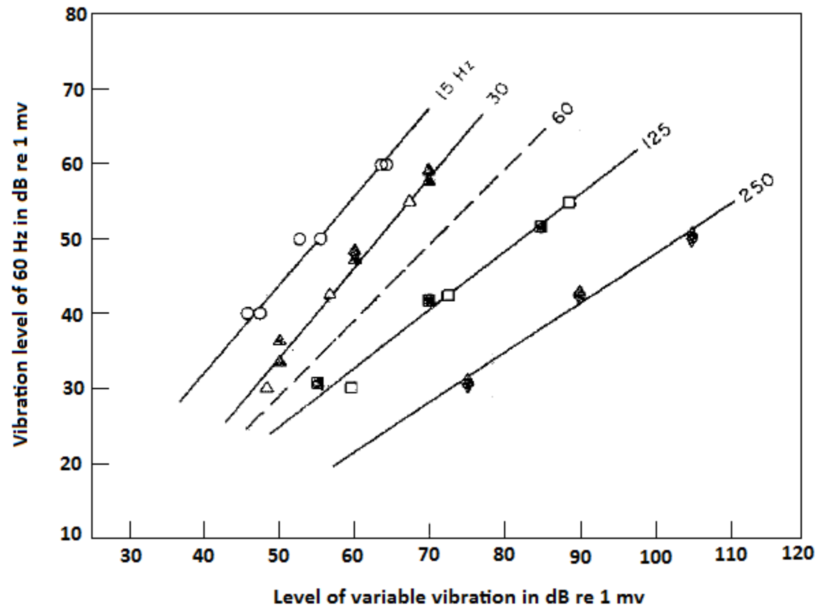


Figure A-14: Matching functions between a 60 Hz vibration and other vibration frequencies. Figure taken from Stevens (1968, p. 224).

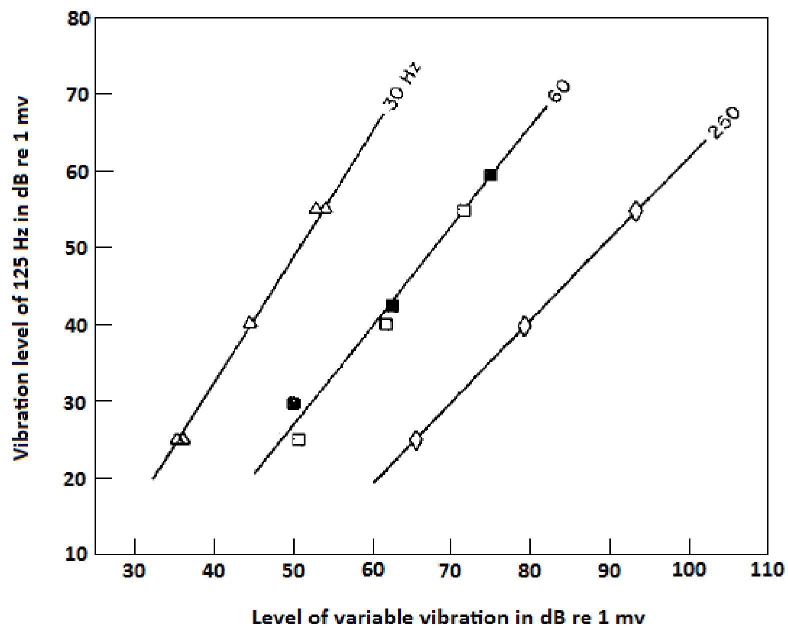


Figure A-15: Matching functions between a 125 Hz vibration and three other vibration frequencies. Figure taken from Stevens (1968, p. 225).

2. Matching by tracking: In this method, subjects were presented with two vibratory stimuli. The level of the reference stimulus was slowly increased. The subjects were instructed to track the intensity of the reference stimulus. This task was done by pressing a button whenever the variable stimulus seemed less intense than the reference stimulus, and releases the button whenever it seemed more intense. Figure A-16 shows examples of tracking recorded for one of the participants.

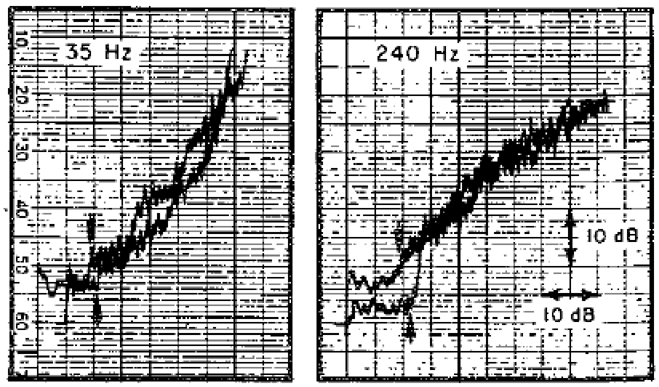


Figure A-16: sample tracking records for one of the subjects. The participant tried to track the intensity of a 100 Hz vibration (Reference vibration) with a variable stimulus at another frequency. Figure taken from Stevens (1968, p.226).

The results of the experiment revealed that the equal sensation functions of vibrations are power functions.

Conclusions:

Perceived intensity of a vibratory stimulus at a given frequency grows as a power function of stimulus amplitude.

Reference:

Summers, I. R., Cooper, P. G., Wright, P., Gratton, D. A., Milnes, P., & Brown, B. H. (1997). Information from time-varying vibrotactile stimuli. The *Journal of the Acoustical Society of America*, 102(6), 3686-3696.

Overview:

Experiments were done to investigate the perception of step changes in stimulus frequency.

Methodology and results:

Vibratory stimuli were presented to the distal pad of the right index finger. The stimuli were periodic signals of 80, 160, 240 and 320ms durations with one octave step change of frequency at their halfway point. For example a signal of 240ms duration was increased/decreased one octave in its frequency after 120ms from its onset. There were also constant stimuli with no step change. Three different waveform types were used for this experiment: Sinewave, monophasic pulse and tetraphasic pulse. Figure A-17 illustrates the waveforms. Vibrations were presented at two different sensation levels, 24 dBSL and 36 dBSL. Participants were almost always able to correctly detect constant stimuli. But there was some unsuccessful discrimination of stimuli with increasing or decreasing frequency. According to the overall results of this experiment which is illustrated in Figure A-18, there was confusion in discrimination of increasing or decreasing frequency.

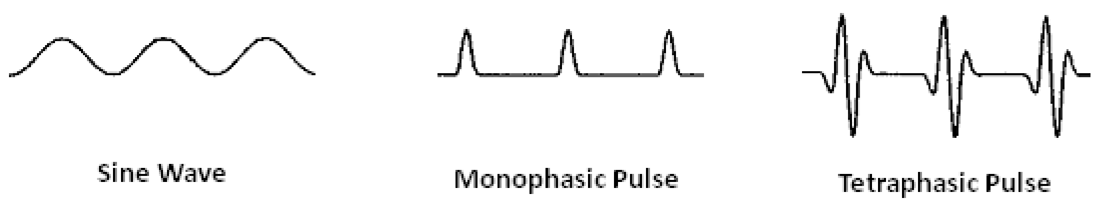


Figure A-17: Three types of waveforms used in (Summers et al., 1997) experiment

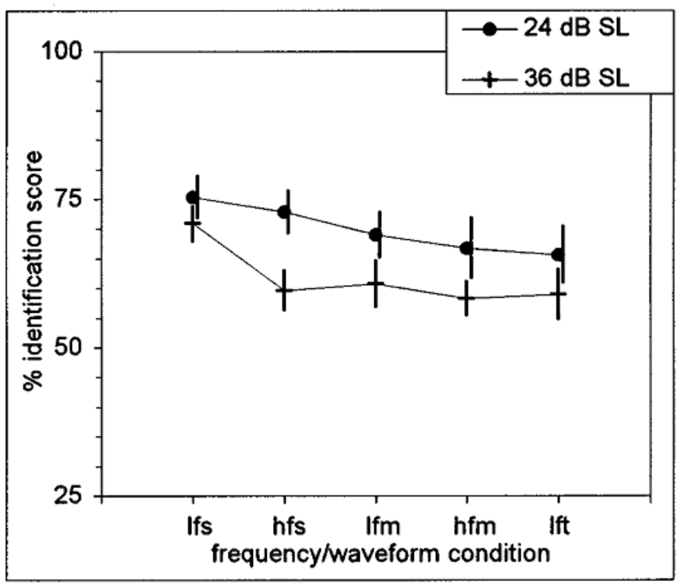


Figure A-18: Overall results of Summer et al. (1997) experiment. Ifs = 50/100Hz sine; hfs = 200/400 Hz sine; lfm = 50/100 Hz monophasic; hfm = 20/400 Hz monophasic; lft = 50/100 Hz tetraphasic. Figure taken from Summers et al. (1997, p.3690).

<u>Conclusions:</u>

Due to uncertainties in change of frequency perception reported in this paper, it is unclear that frequency of vibration would be a useful parameter to be controlled in order to present messages in a vibrotactile display.

Reference:

Van Erp, J. B. F. (2005). Presenting directions with a vibrotactile torso display. *Ergonomics*, 48(3), 302-313.

Overview:

Vibratory stimuli were presented to a group of subjects through a tactor belt. The subjects were asked to indicate the perceived location of vibration by means of the specific apparatus which was provided for this experiment. The response patterns of the subjects are explained and reported in this paper.

Participants wore a tactor belt consisting of 15 vibrotactors embedded equidistantly around the belt's circumference. They sat on a stool which was located in the centre of a circular gap in a horizontally positioned square board. The board level was just above the navel of the participant. One stimulus, consisting of a vibrating tactor, was activated in each trial. The participants were asked to indicate the location of the vibration on a horizontally positioned square board, which they were seated within.

Results:

Considering Figure A-19, the results of this experiment demonstrated that there was a bias between the actual location of the tactors on the torso and the indicated locations by the participants as their response. The bias was toward the midsagittal plane, that is, perceived locations were toward the navel for the tactors located on the abdomen and toward the spine for the tactors located on the back. This result is consistent with the findings of Cholewiak et al. (Cholewiak et al., 2004) and supports the fact that the navel and the spine can be considered as the anchor points of the torso.

Also, all participants showed a pattern in which the lines from the indicated location of the tactor on the square board to the actual tactor spot on the observer's body surface seemed to cross on two points. One of these points exists for the left and one for the right body half, with a mean lateral distance of 6.0 cm between them. This means that observers do not use the body midaxis as the origin for the observed direction.

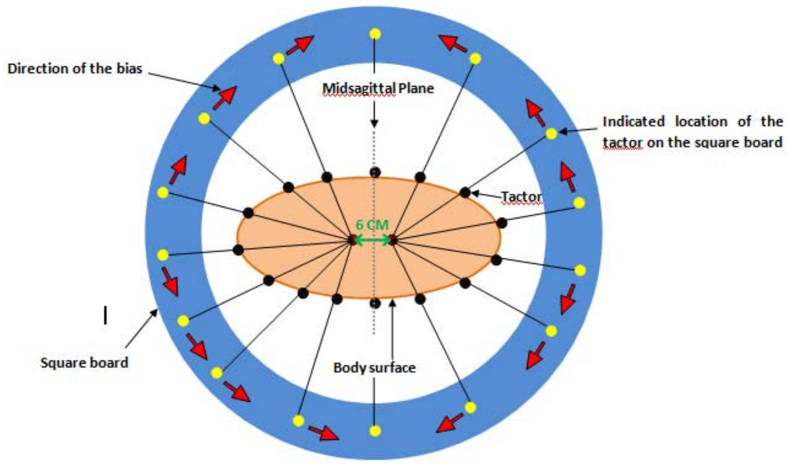


Figure A-19: Schematic top view of the (Van Erp, 2005) experiment and the results.

Conclusions:

The navel and the spine can be considered anchor points of the torso. There are two internal reference points in human body, one for each body half (left and right), and observers do not use the body midaxis as the origin for the observed direction. This suggests that spatial tactile signals should be designed from the internal reference points in the body, and not simply from the midsaggital plane as this reflects how people will tend to interpret the signals.

Reference:

Van Erp, J. B. F. (2005). Vibrotactile spatial acuity on the torso: Effects of location and timing parameters. In *Proceedings of the First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems* (pp. 80-85).

Overview:

Two experiments were executed to investigate the processing of spatio-temporal vibrotactile patterns by the skin of the trunk.

Methodology:

In the first part of the experiment the spatial resolution of vibrotactile stimuli on different locations of the torso was investigated. This was done by placing vertical and horizontal arrays of tactors on the skin of the back and the abdomen. Each presentation consisted of the sequential activation of two vibrotactors. The experimental task was to indicate whether the second tactor was presented to the left or to the right of the first tactor for the horizontal arrays, and above or below of the first tactor for the vertical arrays.

In the second part of the experiment, the effects of the Burst Duration (BD) and Stimulus Onset Asynchrony (*SOA*) on localization performance were assessed. Four pairs of vibrotactors were attached to the back of participants. The center-to-center distance between two tactors within a pair was 2.5 cm. The distance between two pairs was 3.5 cm. The pairs of the tactors were positioned on the back of participants such that their centers were located at -9, -3, +3 and +9 cm with respect to the subject's midline. Figure A-20 shows the arrangement of the tactors for this part of the experiment. Each presentation consisted of the sequential activation of two tactors with 25 combinations of a given *BD*s and *SOA*s. The task of the observers remained the same (indicate whether the second tactor was to the left or to the right of the first tactor).

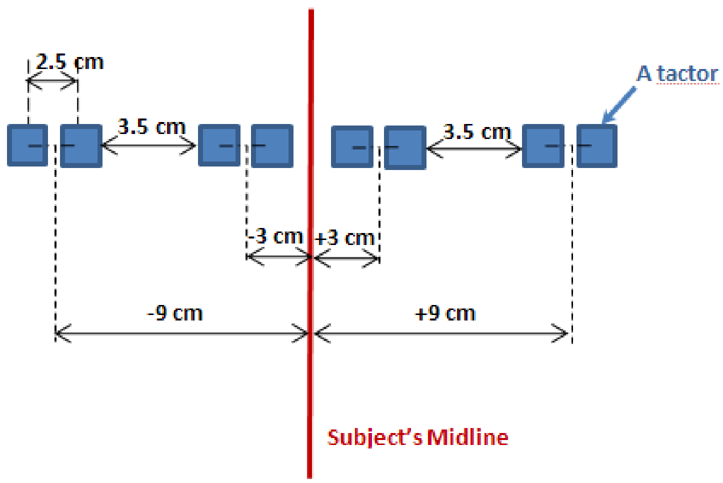


Figure A-20:The tactor arrangement in the second part of the (Van Erp, 2005) experiment

<u>Results:</u>

The results of the first part of the experiment demonstrated a uniform acuity about 2-3 cm across the trunk and there were no acuity differences between horizontally and vertically located arrays. The acuity was better for horizontally oriented arrays located on the spine and the navel and was about 1 cm for these regions. This midline accuracy confirms the fact that the spine and the navel can serve as anatomical anchor points (Cholewiak et al., 2004; Van Erp, 2005), not just because they are anatomical reference points, but because acuity may also be more accurate in these locations.

The results of the second part of the experiment are depicted in Figure A-21. As this figure shows, both *BD* and *SOA* affected the localization performance. Accuracy improved when *BD* and *SOA* increased, and *SOA* was found to have larger effects on accuracy than *BD*. Therefore, there is a trade-off between the speed of stimulus presentation and spatial acuity.

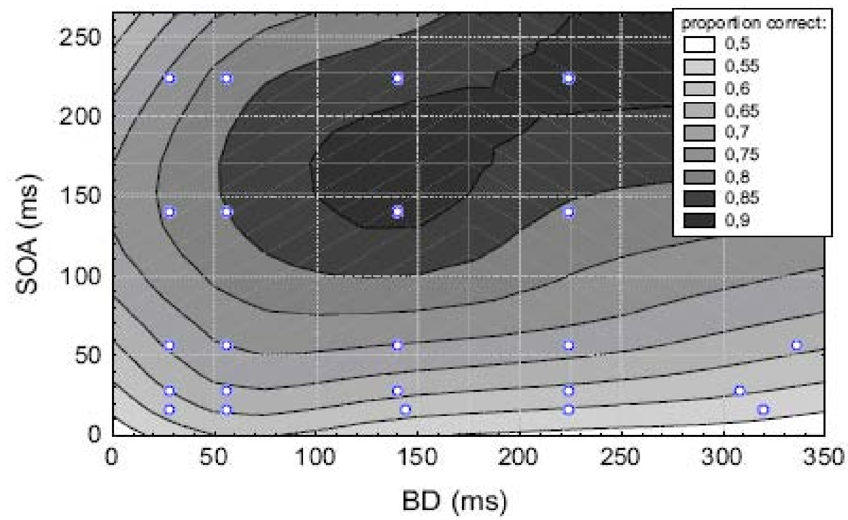


Figure A-21: Effects of the timing parameters on localization performance. "Proportion correct" as function of BD and SOA. Darker colors indicate better performance. Figure taken from Van Erp (2005, p.4).

Conclusions:

Applications which utilize tactile displays and need high spatial acuity can profit from longer *BDs* and *SOAs*. The spatial acuity for vibratory stimuli is relatively uniform over the trunk and it is approximately 3 cm. This acuity is better for horizontally oriented arrays located on the spine and the navel and is about 1 cm for these regions. Localization performance on the skin of the torso improves when *BD* and *SOA* increase.

Reference:

Van Erp, J. B. F., Groen, E. L., Bos, J. E., & Van Veen, H. A. H. C. (2006). A tactile cockpit instrument supports the control of self-motion during spatial disorientation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 48(2), 219-228.

Overview:

The effectiveness of a vibrotactile torso display as a countermeasure to spatial disorientation was investigated in this study.

Methodology:

Subjects wore a vibrotactile display vest consisted of 24 columns of 2 vibrotactors and were seated on a rotating chair. This vibrotactile display was designed to help participants to recover from spatial disorientation condition. The spatial disorientation condition (Pre-SD phase) were simulated by rotating the chair with a constant acceleration for 24 seconds and immediately thereafter, bringing the chair to a standstill condition within 1.2 seconds. The recovery phase was started after 0.5 seconds of standstill condition and during this stage an angular velocity disturbance was presented to the chair. The task of the subjects was to annul the chair's velocity during the recovery phase. This could be done by means of a control knob.

The yaw rotation was represented by sequentially activating the columns of the vibrotactaors around the observer's torso in the horizontal plane. The inside-out and the outside-in coding principals were applied in this experiment. For the inside-out coding of yaw rotation, the vibrotactile signal rotated in the opposite direction of the pilot's rotation (for example the vibrotactile signal rotated clockwise when the pilot rotated counter-clockwise). In the outside-in coding the vibrotactile signal rotated in the same direction of the pilot's rotation. The subject's view was blocked during the experiment.

All of the instrumentations were controlled by a computer. The computer generated the chair velocity signal for creating the spatial disorientation condition and the velocity disturbance signal which was presented to the chair during the recovery phase. The computer also recorded the following information during the experiment:

- Chair velocity control signal
- Chair position
- Chair velocity
- Activated tactile orientation
- Knob position

Two performance measures were computed:

- Recovery performance: This was calculated as the number of spins made by the participants. The number of spins could show the inability of subjects to recover from spatial disorientation.

- Control Performance: This was calculated as the correlation between the disturbance signal and the control input. The result could indicate the capability of participants in counteracting the disturbance

The experiment was designed to investigate two main questions:

1- Does a vibrotactile display help the operators in recovering from spatial disorientation?
2- Is it beneficial to activate the vibrotactile display during pre-SD phase?

Figure A-22: Situation of a subject during the experiment. A subject seated on the rotating chair with a control knob in his hand and the visual cues are blocked. Figure taken from Van Erp et al. (2006, p.221).

<u>Results:</u>

Figure A-23 and Figure A-24 illustrate the results of the experiment for recovery performance and control performance.

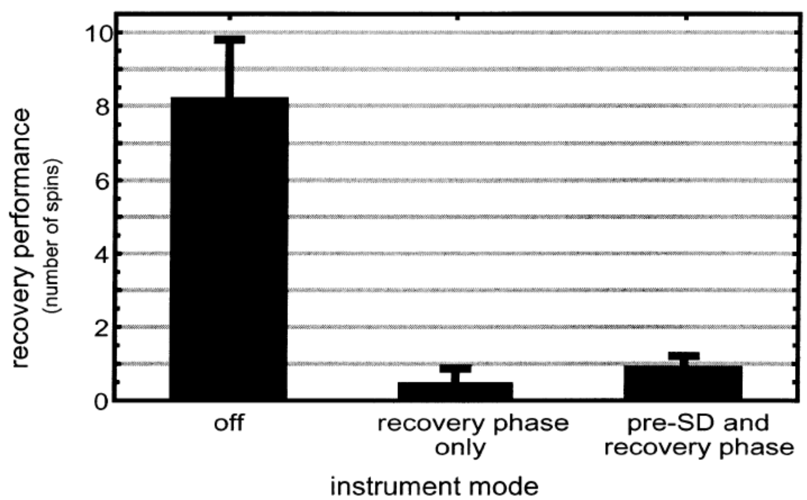


Figure A-23: Effects of using the vibrotactile display on recovery performance of the participants (Lower values indicate better performance). Figure taken from Van Erp et al. (2006, p.224).

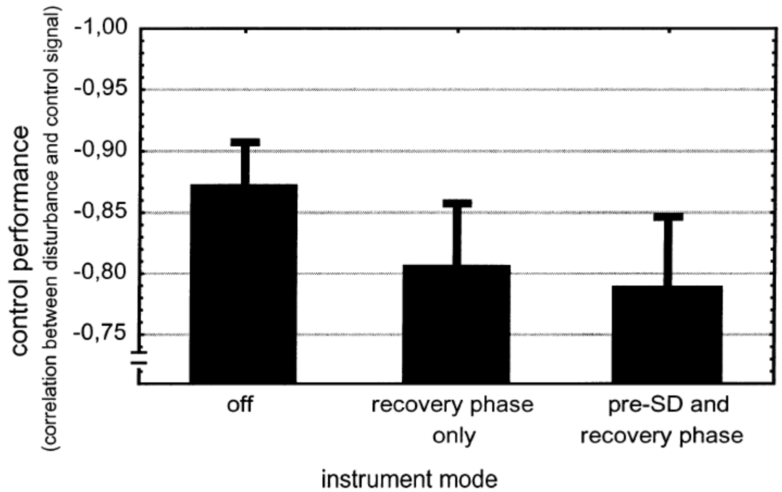


Figure A-24: Effects of using the vibrotactile display on control performance of the participants (Values closer to -1.00 show better performance). Figure taken from Van Erp et al. (2006, p.224).

In regards of the first question, it is obvious from Figure A-23 that a vibrotactile display can support operators in recovering from spatial disorientation.

In regards of the second question, the results of the experiment indicated in Figure A-23 also demonstrated that there is no need to have the vibrotactile display running during the pre-SD phase.

Besides the recovery performance, control performance was also calculated in this study. The task of the subjects was to annul the disturbance during the recovery phase. The results shown in Figure A-24 show that the vibrotactile display degraded the control performance of the subjects.

Conclusions:

A vibrotactile display can support operators in recovering from spatial disorientation.

Reference:

Verrillo, R.T., Gescheider, G.A.(1983). Vibrotactile masking: Effects of one- and two-site stimulation. Perception and Psychophysics.33.379-387

Overview:

Experiments were executed to investigate vibrotactile spatial masking. Spatial masking occurs when two stimuli are presented to two distant locations at different or overlapping times.

Methodology and results:

In this study vibrations were presented to the distal pad of the index finger and the center of the thenar eminence of the right hand. Two different frequencies were chosen for vibrations. One for stimulating *Pacinian corpuscles* (300 Hz) and one for stimulating non-Pacinian corpuscles (13 Hz). The masker intensities were set at -10, 0, 10, 20, 30, 40, and 50 dBSL. Duration for masker and target stimuli were 700 ms and 300 ms respectively. Target stimuli were presented such that they were centered within the masker stimuli. The experiment was executed in three sections.

In the first part of the experiment, both masking and target stimuli were presented to same site (Both of them were presented to the distal pad of the index finger or the thenar eminence of the hand) and subjects were instructed to track the threshold of the target pattern. Both target and masker patterns were 300 Hz stimuli. Figure A-25 shows the results of the first part of the experiment. According to this figure, the amount of masking increases as masker intensity goes above 10 dBSL. Similar results were obtained for the finger and the thenar eminence of the hand.

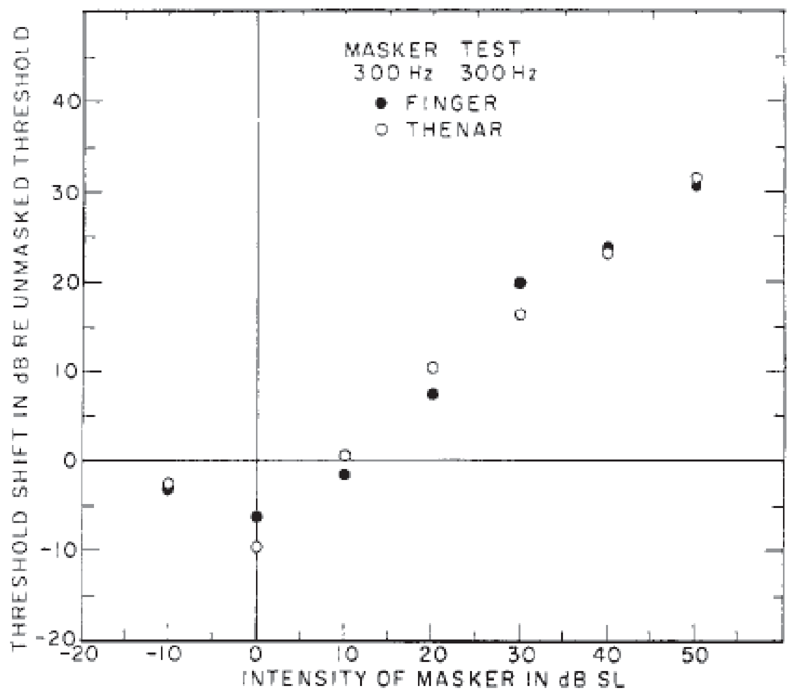


Figure A-25: Vibrotactile threshold shift as a function of the intensity of the masker when target and masking stimuli are presented to the same site. 300 HZ vibrations were used as target and masker stimuli. Figure taken from Verrillo and Gescheider (1983, p.381)

In the second part of the experiment, both masking and target stimuli were presented to the distal pad of the index finger. The masker and the test stimuli were presented in four different ways:

1- 300 Hz masker and 300 Hz target
2- 13 Hz masker and 13 Hz target
3- 13 Hz masker and 300 Hz target
4- 300 Hz masker and 13 Hz target

Considering Figure A-26 which depicts the results of the second part of the experiment, we can conclude that strong masking occurs when vibrations have same frequency and they stimulate same receptor system.

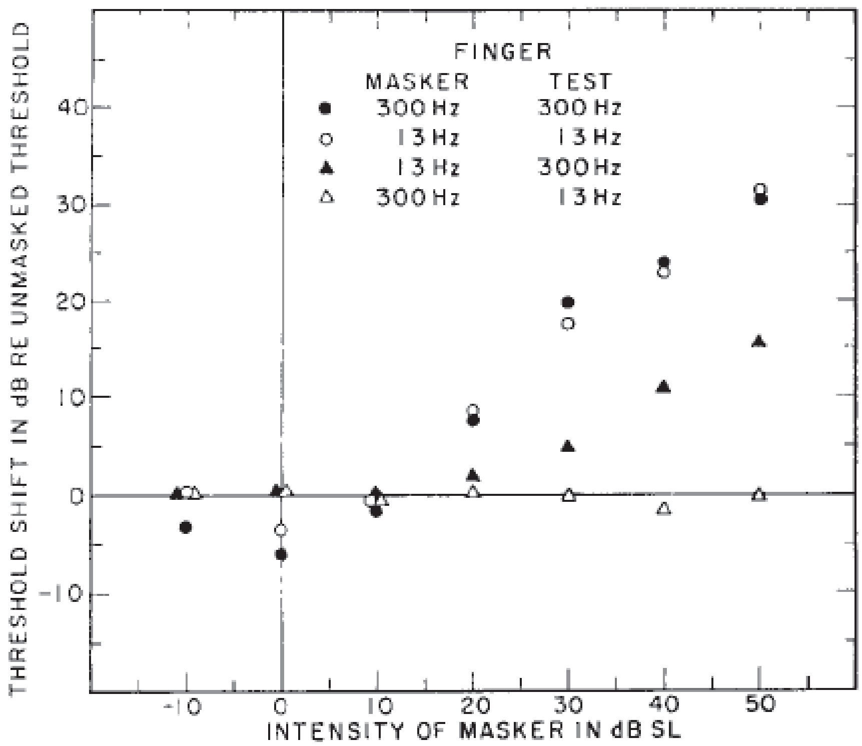


Figure A-26: Vibrotactile threshold shift as a function of intensity of the masker when target and stimuli are presented to the same site. Results of four different combinations of stimuli frequencies are depicted. Figure taken from Verrillo and Gescheider (1983, p.38)

In the third part of the experiment, the target and the masker stimuli were presented to two different sites in order to investigate the effects of remote masking. The masker pattern was presented to the thenar eminence of the right hand and the target pattern was presented to the distal pad of the right index finger. The masker and the target stimuli were presented in four different ways. Similar to those presented in the second part of the experiment. For this part of the experiment subjects were instructed to track the threshold at the distal pad of the index finger. Results of this portion of the experiment are shown in Figure A-27.

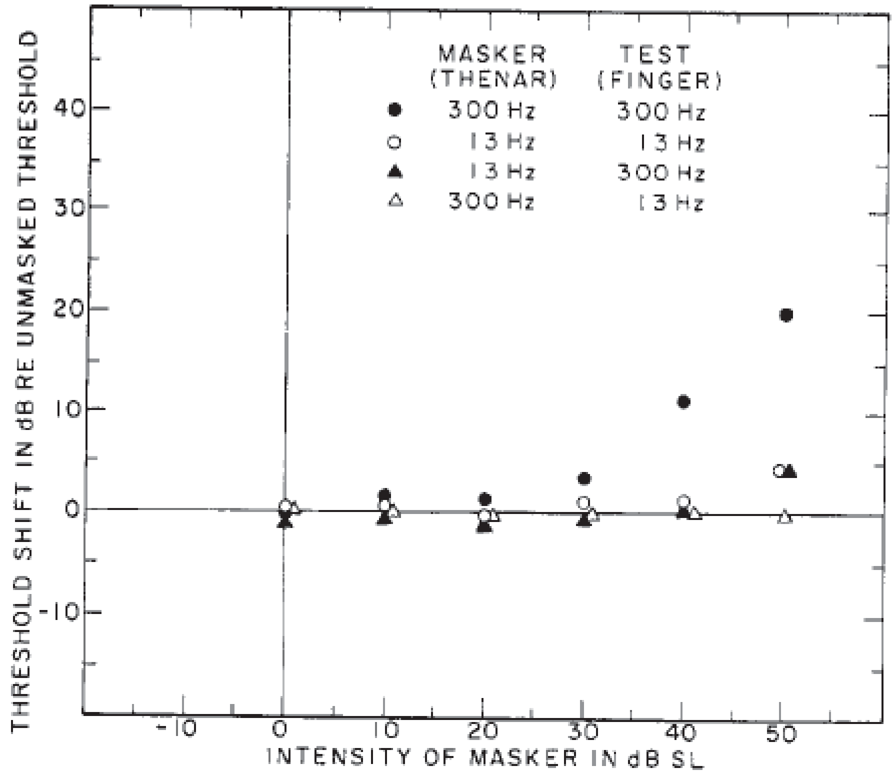


Figure A-27: Vibrotactile threshold shift as a function of intensity of the masker when target and stimuli are presented to two different sites. Results of four different combinations of stimuli frequencies are depicted. Figure taken from Verrillo and Gescheider (1983)

Referring to Figure A-27 we can conclude that remote masking occurs only for high frequency vibrations. Therefore spatial masking is more effective within the Pacinian system. Non-Pacinian system does not demonstrate this characteristic.

<u>Conclusions:</u>

Masking effects may have negative influence on perception of tactile patterns. Therefore, we should be aware of masking properties when designing vibrotactile patterns:

1. When stimulating a single location, strong masking occurs when vibrations have same frequency and they stimulate same receptor system.
2. Spatial masking (remote masking) occurs only within the Pacinian system and for high frequency vibrations

Reference:

Verrillo, R. T., Fraioli, A. J., & Smith, R. L. (1969). Sensation magnitude of vibrotactile stimuli. *Perception & Psychophysics*, 6(6A), 366-372.

Overview:

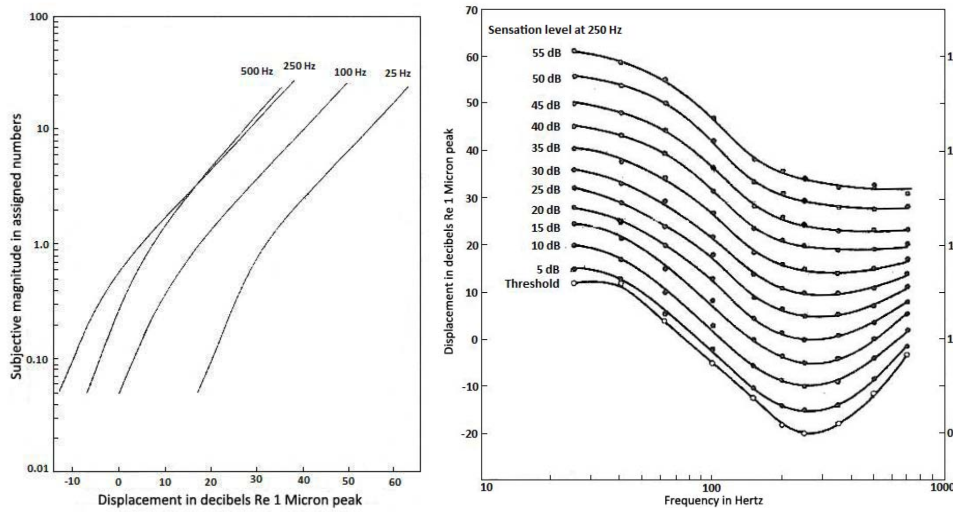
The contours of equal sensation magnitude judgments resulting from the interaction of frequency and amplitude were established in this study.

Methodology:

The stimuli consisted of 10 different vibrotactile frequencies and were presented by a 2.9 cm^2 contactor to the thenar eminence of the right hand. The experiment was done in two main sections. In the first section, a series of 10 stimuli (for each of 10 different vibration frequencies) with different amplitudes were randomly presented. Subjects were instructed to assign numbers regarding to the perceived magnitude of each presented stimulus (magnitude estimation). In the second section, subjects controlled the amplitude of vibrations by means of a plain knob. They were instructed to adjust the amplitude of the vibration such that its magnitude subjectively fit the numbers that had been presented to them (Magnitude Production). For each frequency tested, the geometric mean of the individual responses for magnitude estimation and magnitude production functions was calculated.

Results:

Considering Figure A-28(a), resultant curves indicate that the perceived intensity of vibrations is a power function. The exponents were found to be 0.89 for 25-300 Hz, 0.95 for 500 Hz and 1.2 for 700 Hz vibrations. All of the experimental results were collected and re-plotted in terms of displacement as a function of frequency. The resulting group of curves are presented in *Figure A-28*, illustrating the contours of equal sensation magnitudes. According to these curves, the intensity of a 250 Hz vibrotactile with specific amplitude can be identically perceived as a vibration at another frequency with different amplitude.



(a) (b)

Figure A-28: Subjective magnitudes as a function of absolute displacement (a), Contours of equal sensation magnitudes, the sensation level indications refer to a signal at 250Hz (b). Figures taken from Verrillo et al. (1969, p.370-371).

Conclusions:

Perceived intensity of a vibratory stimulus at a given frequency grows as a power function of stimulus amplitude. Considering Figure A-28, the subjective magnitude of a vibration with a certain frequency can be obtained by means of another vibration with a different frequency, but with slower or higher amplitude. For example, the intensity of a 250 Hz vibrotactile with specific amplitude can be identically perceived as a vibration at 100 Hz frequency with higher amplitude. The results from the mentioned studies reveal the fact that there is a major interaction between frequency and amplitude of a vibrotactile stimulus. Therefore, it is recommended to change only one of these parameters when using vibrotactile displays

Reference:

Verrillo, R. T. (1963). Effect of contactor area on the vibrotactile threshold. *The Journal of Acoustical Society of America*, 35(12), 1962-1966.

Overview:

Sensitivity to vibration on the volar skin of the hand as a function of tactor properties was measured.

Methodology:

A vibrotactor was positioned under a table and its contactor obtruded through a hole which was located on the table top. Adapter rings with different diameters that could be set inside the hole of the table were fabricated. Therefore, it was possible to control the gap between the contactor (the portion of the vibrotactor in contact with the skin) and the edge of the rigid support. Subjects were seated beside the table with their right arm rested comfortably on it. Therefore, it was possible to place the fingers over the hole. The volar surface of the second phalanx on the middle finger and the first metacarpal of the thumb were the regions of testing. The vibrotactor position could be adjusted vertically by a jack to provide different levels of pressure upon the skin. Vibrations were presented in the frequency range 25-640 Hz.

The main goal of the experiment was to find out the contactor properties that control the vibrotactile threshold. Two hypotheses were investigated in this experiment:

1. The contactor area is a significant parameter of vibrotactor stimuli.

2. The gradient or curvature of the skin displacement at the edge of the contactor is a significant parameter of vibrotactor stimuli.

To investigate the accuracy of these hypotheses three types of contactors were used. *Figure A-29* illustrates the cross section of the contactors.

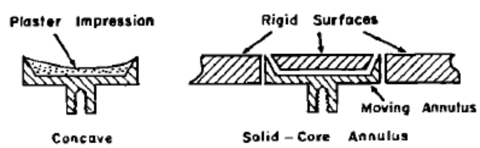


Figure A-29: Cross section of three types of contactors. Figure taken from Verillo (1963, p.1964).

Results:

Figure A-30 illustrates the results of the experiment.

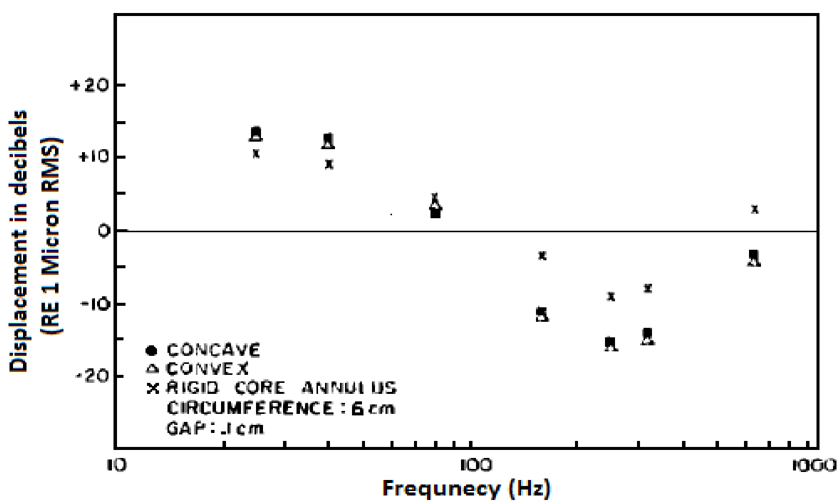


Figure A-30: The vibrotactile thresholds for three types of contactors, all having the same circumference. Figure taken from Verillo (1963, p.1964).

Considering Figure A-30, despite the fact that the gradient or the curvature of the skin is less when using convex contactors, the amount of displacement for detection threshold is approximately identical for concave and convex contactors.

Results of the experiment confirmed the first hypothesis and rejected the second hypothesis. Therefore, the detection threshold of vibratory stimuli is a function of contactor area.

In order to further investigate the hypothesis that the area of the contactor is a controlling parameter of a vibrotactile stimuli, in another experiment, a series of contactors with different areas (0.005, 0.02, 0.08, 0.32, 1.3, 2.9, and 5.1 cm^2) was used. The size of the gap between the contactor and a rigid surface was maintained constant at 1mm by means of adapter rings. Figure A- 31 illustrates the results of this experiment.

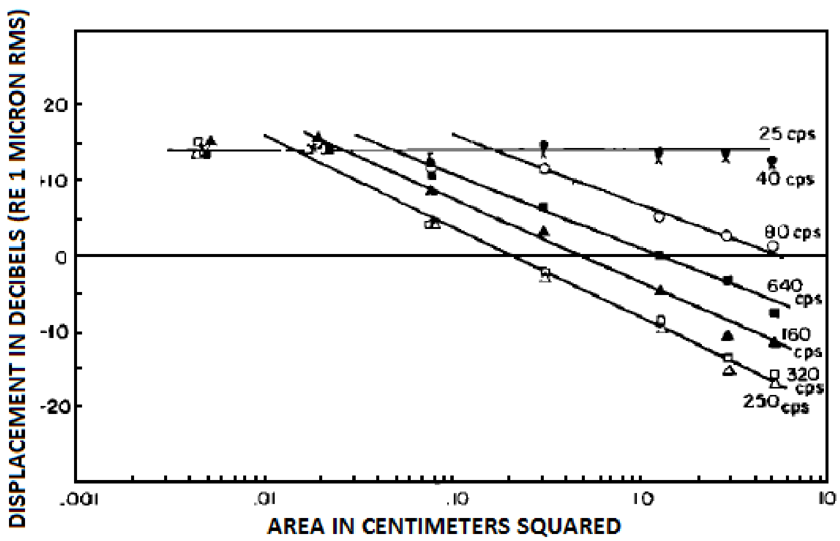


Figure A- 31: The vibrotactile threshold as a function of contactor area. Figure taken from Verillo (1936, p. 1964).

When the size of the gap between the contactor and the rigid surface is controlled and maintained constant at 1 mm, the area of the contactor emerges as a controlling parameter of vibrotactile stimuli.

Conclusions:

The area of the vibrotactor's contactor is a controlling parameter in vibrotactile detection threshold when the contactor is surrounded by a rigid surface. Vibrotactile detection threshold decreases as the contactor area increases.

Reference:

Verrillo, R. T. (1962). Investigation of some parameters of the cutaneous threshold for vibration. *The Journal of the Acoustical Society of America*, 34(11), 1768-1773.

Overview:

Sensitivity to vibration on the glabrous skin of the hand as a function of frequency and tactor properties was investigated in a study.

Methodology:

A vibrotactor was positioned under a table and its contactor obtruded through a hole which was located on the table top. Adapter rings with different diameters that could be set inside the hole of the table were fabricated. Therefore, it was possible to control the gap between the contactor (the portion of the vibrotactor in contact with the skin) and the edge of the rigid support. Subjects were seated beside the table with their right arm rested comfortably on it. Therefore, it was possible to place the fingers over the hole. The volar surface of the second phalanx on the middle finger and the first metacarpal of the thumb were the regions of testing. The vibrotactor position could be adjusted vertically by a jack to provide different levels of pressure upon the skin. Vibrations were presented in the frequency range 25-640 Hz.

Results:

Figure A-32 shows the results of the experiment. Considering this figure, the detection threshold of vibrotactile stimuli as a function of frequency was found to be a U-shaped curve which has its minimum in the region of 250Hz.

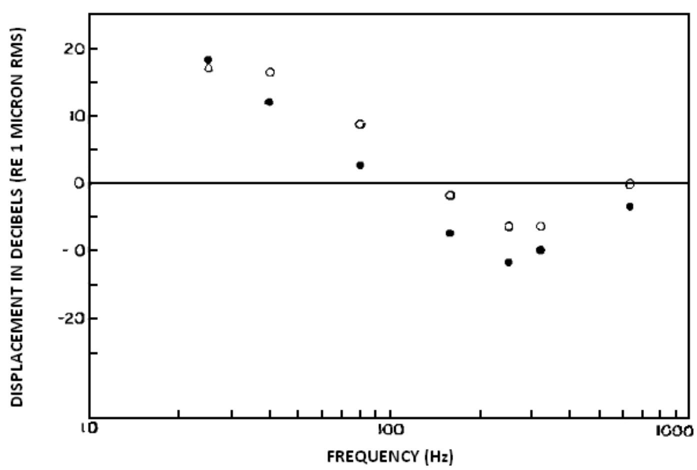


Figure A-32: Detection threshold of vibration on two regions of the hand. Middle of the first metacarpal of the thumb (open circles) and the volar surface of the second phalanx on the middle finger (closed circles). Contactor area 0.283 cm2. Figure taken from Verillo (1962, p. 1770).

Considering Figure A-33, the results of the experiment also revealed that the direction threshold of vibration decreased as the contactor was pressed further into the skin.

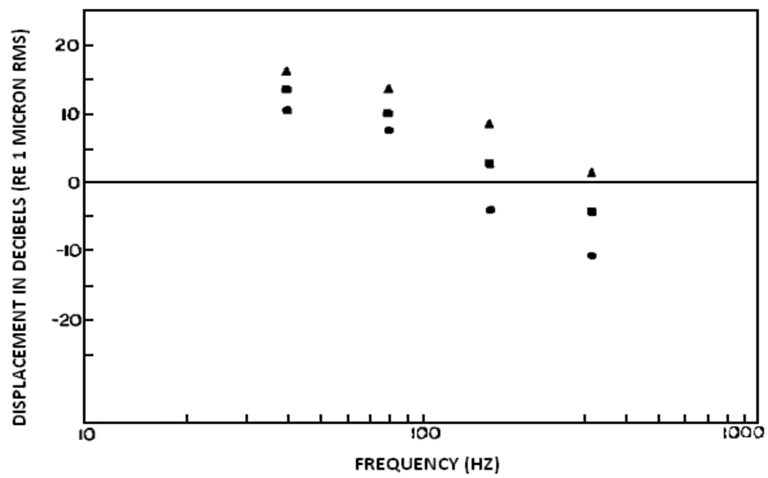


Figure A-33: Comparison of the threshold for vibration at three contactor heights.).5 mm below table surface (triangles), 0.5 mm above table surface (squares) and 1.5 mm above table surface (circles). Contactor area 0.113 cm^2. Vibrations were presented to the finger. Figure taken from Verrillo (1962, p. 1770).

Conclusions:

The detection threshold as a function of frequency for the volar surface of the fingers is a U-shaped curve which has its minimum in the region of 250Hz. Therefore, when using vibrations to present information through a vibrotactile display, vibratory stimuli should have 250 Hz frequency. Detection threshold decreases as the contactor, is pressed further into the skin. Therefore, when information is being presented through a vibrotactile display, the performance can be improved by pressing contactors further to the skin (to provide better contact with the skin)

Reference:

Wilska, A. (1954). On the vibrational sensitivity in different regions of the body surface. *Acta Physiologica Scandinavica*, 31(2-3), 285-289.

Overview:

The minimum amplitude for detecting 25-1280 Hz vibratory stimuli was measured over different locations on the body (detection threshold).

Methodology:

Vibrations in the frequency range of 25-1280 Hz were presented to the different locations of the body. Detection threshold of vibrations were measured. The frequencies used were 25, 45, 77, 125, 200, 270, 360, 450, 580, 750, 860, 1020, 1150 and 1280 Hz. The contactor of the vibrotactor was a cylindrical piece of wood with 1 sq cm in area. <u>Results:</u> It was found that the lowest threshold amplitudes are within the frequency range 200-450 Hz. At 200Hz vibrations, the finger tips have the smallest threshold of 0.07 μm, whereas in the abdominal and gluteal regions this number increases to a maximum of 14 μm. Over the entire frequency range of the experiment (25-1250 Hz), hands were found to be the most sensitive while abdominal and gluteal regions were found to be the least sensitive regions of the body.
<u>Conclusions:</u> The lowest sensory threshold amplitudes for vibratory stimuli detection are within 200-450 Hz. Hands are the most sensitive and abdominal and gluteal regions are the least sensitive regions of the body.

<u>Reference:</u> Yanagida, Y., Kakita, M., Lindeman, R.W., Kume, Y., & Tetsutani, N. (2004). Vibrotactile letter reading using a low-resolution tactor array. In *Proceedings of the 12th International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems* (pp. 400-406).
<u>Overview:</u> The ability of subjects in recognizing vibrotactile patterns which were used to present English letters and numbers to their lower back was investigated. <u>Methodology:</u> The patterns were presented through a 3×3 tactor array affixed to the backrest of an office chair. The sequential presentations of the patterns were such that they were tracing the trajectory in the same order as hand writing. Vibrotactile patterns for 10 digit numbers and all 26 capital alphabet letters were presented to the subjects. The presentation sequence for “O” and “0” (zero) and “Z” and “2” were identical. Therefore, 34 vibrotactile patterns were generated. Figure A-34 illustrates the some examples for the activation sequence.

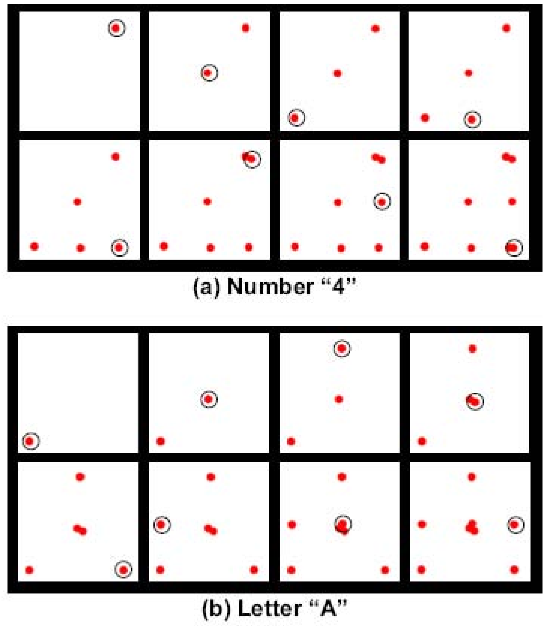


Figure A-34: Sequential activation of vibrotactors to present the number 4 (a) and the letter A (b). Figure taken from Yanagida et al. (2004, p.4).

Burst duration and inter-stimulus interval were 500 ms.

Results:

Among numeric letters, participants could recognize number "1" with the accuracy of 100%, followed by "8","5","6","9","4", and "2". They could correctly recognize number "3" for 77.8% of the time. For the alphabet letters subjects were able to recognize "E", "O", "Q", and "T" with the accuracy of 100% and they could recognize "S" for 62.1% of the times. (the number "3" and the letter "S" were the least recognizable patterns)

The overall ratio of 87% correct letter or number recognition was recorded for this experiment.

Conclusions:

The results of the experiments demonstrated that vibrotactile spatio-temporal patterns presented to the torso can be recognized with high accuracy. Therefore, these patterns can be considered as a reliable option to present information to operators through vibrotactile displays.

A.3 Auditory Display Design and Urgency

Reference:

Ho, C., Nikolic, M. I., & Sarter, N. B. (2001). Multimodal information presentation in support of timesharing and effective interruption management. In *Proceedings of the 20th Digital Avionics*

Systems Conference (pp.5D2/1-5D2/8).

Overview:

This paper examined various methods to support interruption task management by distributing tasks across different modalities and manipulating the amount of information available to the subject about the pending task. The purpose of the study was to explore effective ways of presenting operators with urgency information to support interruption task management. Subjects were required to perform air traffic control tasks (visually) in which interruption tasks were presented and could either be completed through the visual, auditory or tactile modality. When a red box flashed on a visual display, subjects were asked to push a button in which the interruption task would be presented. This task involved counting a subset of cues presented in one of the modalities mentioned earlier. The visual interruption task consisted of flashing circles, the auditory task consisted of slow and fast patterns of "beeping sounds" and the tactile task consisted of vibrations presented to the subjects' right and left inner wrists.

This study consisted of two groups: (1) abridge group in which the subjects in this group were presented with information in regards to the interruption task in terms of urgency, time required to complete the task, and modality of the task and (2) basic group in which subjects were only informed about the presence of a pending task. Overall results demonstrated that presenting subjects with information about the nature of the pending interruption task, "helped participants to schedule and manage interruptions more effectively." This paper also cites research that has demonstrated that different types of information (e.g. source of interruption, task urgency, task completion duration, and task modality) are useful sources to assist the operator with task prioritization, effective scheduling and minimal crossmodal interference. Another interesting result involved modality preferences of the interruption tasks; subjects preferred the auditory modality than the tactile modality, followed by the visual modality. 31 out of the 32 participants in this experiment reported the visual interruption task as the most difficult task to perform. A possible explanation for this finding could be explained through the *multiple resource theory*. Since the primary task (air traffic control task) was presented visually therefore the visual resource pool is already being exercised thus participants may be attempting to avoid intramodal interference.

Conclusions:

Presenting information about the nature of an interruption task can significantly improve the operator's performance by assisting with task management. This is especially true when the task has a high urgency level; operators will be more likely to attend to the high urgency level task faster. Thus in terms of the project's overall goals, if operators will be required to attend to interruption tasks, or even multiple tasks simultaneously, it is vital for the operator to have access to additional information in terms of urgency, task duration etc.

Reference:

McNeer, R., Bohórquez, J., Ozdamar, O., Varon, A., & Barach, P. (2007). A new paradigm for the design of audible alarms that convey urgency information. *Journal of Clinical Monitoring and Computing*, 21(6), 353-363.

Overview:

In this study auditory alarms with different structures were presented to the subjects and the judgments of the subjects regarding the perceived urgency level of these sounds were recorded. Three groups of sounds were designed for experimental purposes in this study: Harmonic interval sounds, Melodic interval sounds and Duty cycle sounds. Visual representation of each of these groups are illustrated in Figure A-35.

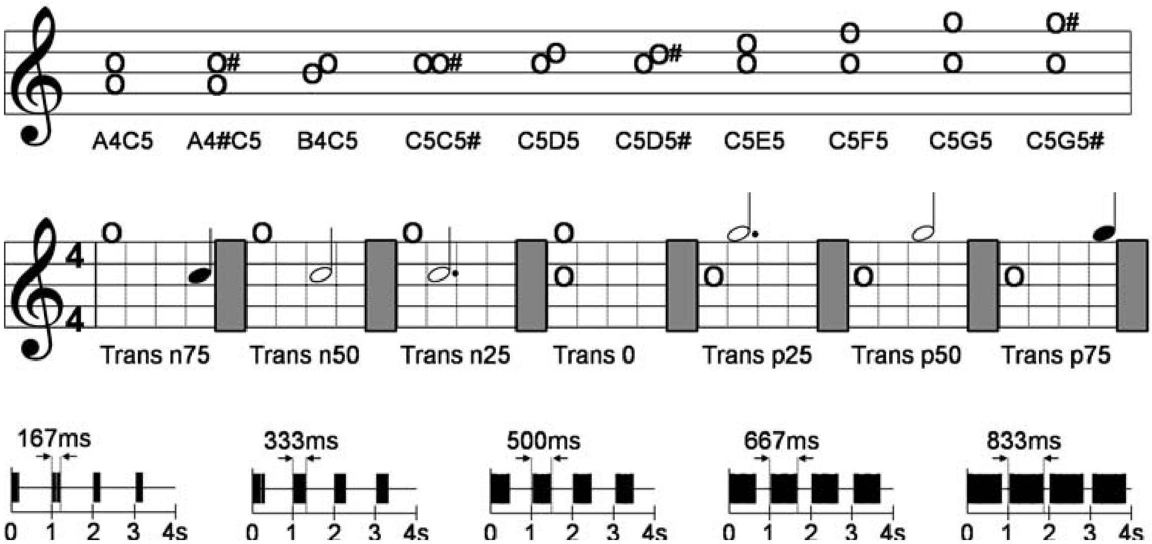


Figure A-35: representation of three groups of sounds in McNeer et al. experiment

Referring to Figure A-35, First panel represents the harmonic interval sounds consisted of ten two-tone chords. The second panel illustrates the melodic interval sounds consisted of seven two-tone chords. The two musical notes at each chord have different onset time relative to the other note. The third panel shows the duty cycle sounds consisted of four presentations of a tone in a 4 sec period with different pulse widths. Each of the auditory alarms were presented to the subjects and they were instructed to rate the level of perceived urgency level by assigning a number between 1-100.

The final results of this experiment are depicted in Figure A-36. As can be seen from this figure, the harmonic interval sounds covered the greatest range of perceived urgency levels (35 -80%). The range of urgency was smallest for the melodic interval sounds (52-72%) and finally the urgency levels for the duty cycle sounds ranged from 38% to 70%.

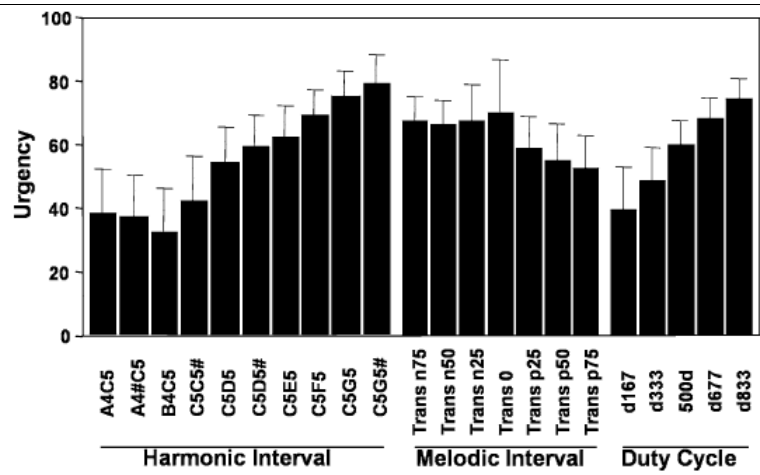


Figure A-36: Perceived urgency levels of sounds in McNeer et al experiment

Conclusions:

The harmonic interval sounds cover relatively good range of perceived urgency levels (35 -80%) and would perform better than the melodic Interval sounds and duty cycle sounds in presentation of different levels of urgency in auditory alarms.

A.4 Crossmodal Attention

Reference:

Beierholm, U. R., Kording, K. P., Shams, L., & Ma, W. J. (2007). Comparing Bayesian models for multisensory cue combination without mandatory integration. In *Proceedings of the 21st Annual Conference on Neural Information Processing Systems (NIPS 2007)*.

Overview:

This paper reviews and compares several Bayesian models of multisensory perception (Maximum-Likelihood Estimation, Cue Integration with Consideration of Prior Knowledge, and Casual Inference Model), as well as evaluates the Bayesian models against a psychophysics experiment. The psychophysics experiment tested participant performance in an auditory-visual spatial localization task, where the integration of modalities was not required.

Previous research in the use of Bayesian models had focused on determining the source and cause of each cue. However, this paper focused on how Bayesian modeling could be used for resolving conflicting information between different sources through cue integration. Despite a large amount of experimental data, no general theory exists which is able to explain multisensory perception across a wide range of cue conflicts. Beierholm reasoned that the casual inference model would be most appropriate for modeling the integration of conflicting cue information.

To evaluate this hypothesis that the casual inference model was most appropriate, an experiment was conducted where subjects were presented with a short visual and auditory stimulus at the same time. This stimulus could be located anywhere in one of five locations on an imaginary horizontal line. The subjects had to report using a key press the perceived position of the auditory and visual stimulus. The response distributions were obtained for the three models: a traditional cue integration model (maximum-likelihood estimation), a bisensory stimulus prior model, and a causal inference model. From this, Beierholm found that the casual inference model best fit the participant data collected in the experiment.

Conclusions:

Bayesian models provide information on how the brain processes probabilistic sensory information. They provide insight as to how the brain handles both small and large conflicts between incoming stimuli. Bayesian models can serve as an alternative method for analyzing how a human operator would interpret a multimodal interface. Use of a Bayesian model can be more cost effective than running a large experiment to evaluate how a human participant may interpret conflicting pieces of multimodal information in a multimodal interface. Designers may also make use of Bayesian models to ensure that conflicting information in different modalities can be easily resolved (as predicted by the models).

Reference:

Chung, P. H., & Byrne, M. D. (2004). Visual cues to reduce errors in a routine procedural task. In K. Forbus, D. Gentner, & T. Regier. (Eds.), *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (pp. 227-232).

Overview:

This paper attempts to evaluate the effectiveness of "visual cues as error interventions in computer-based routine procedural tasks" through reviewing past research findings. Routine procedural tasks include tasks that occur regularly in a routine such as one pumping gas, photocopying documents, etc. Some important findings that the authors gathered are as follows:

Operators can still make errors within highly familiar tasks. For example, many of us have often forgotten the original copy of a document in a photocopier after making copies or forgot to put the gas cap back on after pumping gas. These tasks are simple procedures people engage in on a regular basis but sometimes fail to complete a step in our overall the overall goal (e.g. pumping gas). A hypothesized explanation for this is that the working memory is experiencing high workload leading to a "goal loss or omission of a step from the current task." This paper defined post-completion errors as “errors that occur when the task structure demands that some action is required after the main goal of the task has been satisfied or completed.” Humans tend to generate errors during post-completion steps (e.g. forgetting the last step of retrieving the original document from the photocopier) within subtasks and larger tasks. The list provided below are some predictions of post-completion errors and characteristics of a successful reminder cue:

- Salient cues (e.g. blinking lights/flashes) are sufficient to prime a post-completion action (to serve as a reminder to complete the post-completion task)
- It should not be necessary to put the post-completion action on the critical path
- Reminders at the beginning of tasks will not help a post-completion task error at the end

due to the reminder being masked by other goals

- “Just-in-time priming from environmental cues are a reliable reminder." Just-in-time cues serve as a reminder cue that are presented to subjects/users when it was necessary to complete the specific task (e.g. an auditory cue such as a beeping noise to remind users to retrieve their bank card from ATM machines after they complete their transaction). In an experiment where participants were given a number of tasks, each with its own subtasks, "just-in-time" cues showed a reduction in post-completion errors.
- Visual cues that are colourful are effective in guiding operators to desired points of activity
- To attract attention on visual displays, movement (e.g. blinking, position change), size and shape differentiation, colour, brightness, texture and surroundings (borders, background colour) are effective. It is important to note that these techniques must be used sparingly due to the fact that users will ignore them if they are used in meaningless situations or in an abundant amount.

Conclusions:

These findings stress that operators can forget a post-completion step even if they are extremely familiar with the task’s procedure. This could result in detrimental performance due to the post-completion task possibly being an important step of the overall task. The guidelines provided above can assist in the design of interfaces to ensure that operator errors are minimized, especially in regards to procedural tasks.

Reference:

Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics*, 16(2), 409-412.

Overview:

Colavita conducted various experiments that suggested humans have a visual sensory dominance.

Experiment 1: Participants were presented with either unimodal auditory, unimodal visual or bimodal (audio and visual) targets in which they were told to respond to these targets by pressing a "light key" (visual response key) if they recognized a visual target or a "tone key" (audio response key) in the case they recognized an auditory target. These targets were presented in a random manner but it was mandated that each stimulus be used on 50% of the trials. Results indicated that when bimodal targets were presented, participants responded to the visual component more frequently than the auditory component. Participants reported that they did not even notice the auditory component of the bimodal targets, exemplifying a prepotency of visual stimuli over the auditory stimuli.

Experiment 2: Colavita then wanted to determine whether this tendency would still occur if the intensity of the auditory stimulus was increased relative to the visual stimulus by a factor of two which was carried out by increasing the intensity of the 4,000- Hz tone “until it was twice as loud as the light was bright (50fc)”. It is important to note that Colavita conducted the second experiment in the same format as the first with the exception of auditory stimuli intensity modification. Results showed that the prepotency of visual stimuli over auditory stimuli still existed to the same degree as the first experiment.

Experiment 3: After the first two experiments, Colavita conducted a third experiment to determine whether the ambient illumination level in the experimental room had caused the results in the first two experiments. He conducted this experiment in the same format as the first experiment with the following three exceptions: (1) all windows in the experiment room were uncovered, (2) room lights were turned on to provide normal illumination, and (3) participants were not provided with a verbal "ready" signal before each trial. Regardless of the manipulations that Colavita conducted, the overall result observed in Experiment 1 was also observed in the subsequent experiments; there was an apparent prepotency of the visual stimulus over the auditory stimulus.

Conclusions:

This suggests that the tendency for humans to be visually dominated must be considered when designing and implementing interfaces. For example, an interface designer must note that when the interface conveys information in a modality other than the visual modality, the user may be more prone to ignore the information and direct their attention towards something in their visual field due to visual dominance.

Reference:

Farah, M. J., Wong, A. B., Monheit, M. A., & Morrowt, L. A. (1989). Parietal lobe mechanisms of spatial attention: modality-specific or supramodal. *Neurophysilogical, 27 (4).*

Overview:

In this fundamental paper, the authors presented the concept of a supramodal attention system. This concept included the theory that a focused attention location may be integrate across sensory modalities. However, the initial theory only considered the attention to one location, and did not consider the ability of humans to divide their attention across sensory modalities.

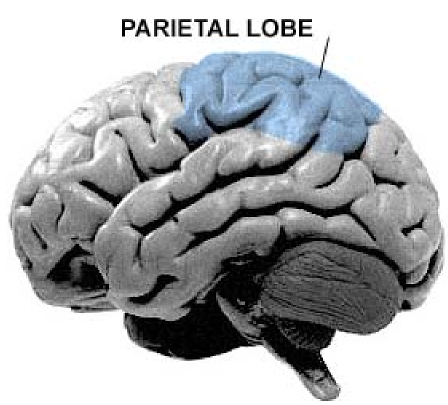


Figure A- 37: Parietal Lobe

In this paper, a study was completed to compare two theories: the theory of a single supramodal attentional system, and the theory that attentional resources are divided into separate, modality-specific subsystems. The study utilized subjects suffering from parietal lobe lesions. The parietal lobe is responsible for integrating sensory information from different modalities. Thus, the effectiveness of stimuli and the division of attentional resources could be studied by comparing responses from the normal side of the brain/body to the side of the brain/body suffering from the lesion.

There were two conditions evaluated. In both situations, the subject was presented with visual stimuli. However, in the first cue condition, this stimuli was preceded by an auditory cue stimuli, and in the second cue condition, the stimuli was preceded by another visual cue stimuli. For both cue situations, subjects were slower to respond to invalidly cued targets occurring on the side of the body opposite of the lesion.

The results from the study showed that there was attentional disengagement impairment for visual targets with auditory cues. Therefore, the parietal lobe's attentional mechanism operates based on the representation in space where both visual and auditory stimuli are represented. This supports the authors' proposed theory of the existence of a supramodal attention system.

Conclusions:

Understanding the body's division of attention resources can help interface designers to understand how modalities can be combined and integrated into information presentation. However, more work needs to be completed in this area, because there are conflicting models regarding the division of attentional resources.

Reference:

Franconeri, S. L., Hollingworth, A., & Simons, D. J. (2005). Do new objects capture attention?. *Psychological Science*, 16(4), 275-281.

Overview:

Although a lot of research suggests that the appearance of a new object captures attention which is called a new-object hypothesis (Hillstrom & Yantis, 1994; Jonides & Yantis, 1988; Jonides & Yantis 1990), recent findings show that luminance-based transients such as motion and brightness can capture attention (called the transient hypothesis). This study investigated whether new objects captured attention because the visual system is sensitive to new objects or because it is sensitive to transient qualities that new object possess. Experiments required subjects to participate in a visual search task were conducted in which subjects were presented with a visual search task. In the experiments conducted, subjects were initially presented with an annulus surrounded with a set of number eight placeholders. The annulus then began to shrink and passed over the annulus over a 180ms interval. In Experiment 1, the placeholders were completely covered for 10ms and in Experiment 2, the annulus did not completely cover the placeholder at any given time. The figure below depicts the different conditions within each experiment of various occlusion conditions

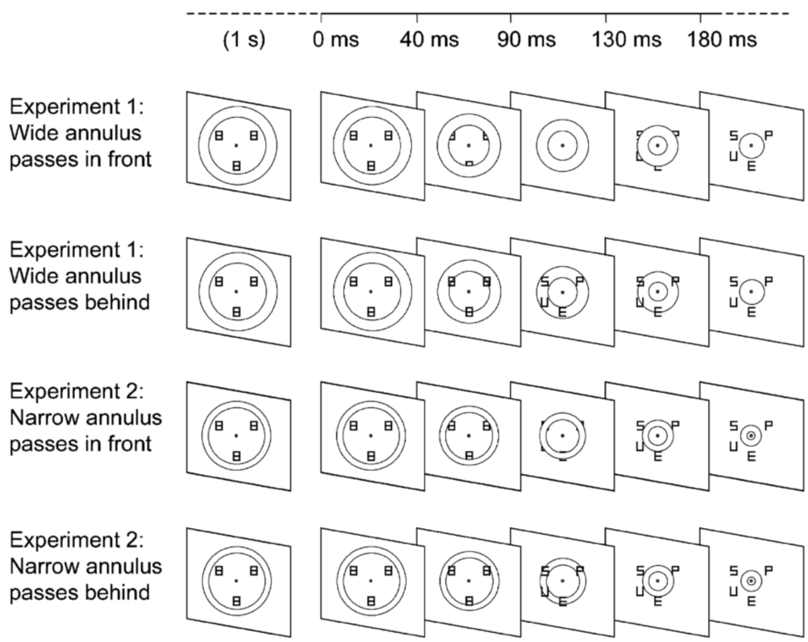


Figure A-38: Occlusion and control conditions in Experiments 1 and 2 (p. 277)

When the annulus completely covered the placeholders, they were replaced with letters (new object) and participants were required to search for a target which was either H or U (the letter target was either a new or an old letter). Franconeri, Hollingworth, and Simons concluded that if the new-object hypothesis was valid, then the new letter should "have been given search priority" in their visual search task. However, the new letters will not be "given search priority" if the luminance transients capture attention. The authors said this is because "the luminance transient produced by the disocclusion of the new letter was equal to the transients created by the disocclusion of the old letters. Both experiments had control conditions which consisted of the annulus passing behind the objects so that subjects could see the "unique onset transient created by the new letter." If the transient hypothesis is valid, then the new letter should only be able to capture attention in the controlled condition.

Results showed new letters were not prioritized in the visual search when it appeared behind the annulus and the accompanied onset transient was not visible even though this letter was a new object. However, the new letter captured subjects' attention when it appeared in front of the annulus and when the accompanied transient was visible. Thus, the authors concluded that new objects did not capture attention unless it possessed a strong luminance-based transient such as motion and looming.

Conclusions:

The evaluation in this study strongly inclines that presenting one with a new object is not sufficient to capture one's attention. However, if the object has a strong luminance-based transient, attention can be more readily and easily captured.

Reference:

Gallace, A., & Spence, C. (2009). The cognitive and neural correlates of tactile memory. *Psychological Bulletin, 135*(3), 380406.

Overview:

This paper presents a review on past research pertaining to the storage and retrieval of data regarding tactile events, referred to as tactile memory systems.

From past work, the authors suggest that tactile memory is divided into several different neuranatomic components; where each component is a function of the properties of the tactile stimulus. In the past, it has been determined that spatial information is stored in the secondary somatosensory cortices and the posterior parietal cortex. However, research has also shown that the haptic information of object requires the engagement of the insula as well. This observation is similar to that for visual and auditory memory, which the memory function is divided into a number of functionally distinct subsystems.

In addition, from the literature review, Gallace and Spence present the concept that tactile memory occurs in the same brain networks which are involved in the initial processing of sensory information. It is clear that the neural components for tactile memory are not just reserved to the tactile modality. Rather, they share connections with the neural networks for perception and memory. This supports the theory of a single supramodal sensory system.

Conclusions:

This paper supports the theory of a single supramodal sensory system for dividing attentional resources, with a specific focus on memory. Understanding the workings of human memory can provide future suggestions for how humans can adapt to past events in an operational environment.

Reference:

Goldstein, I. L., & Dorfman, P. W. (1978). Speed and load stress as determinants of performance in a time sharing task. *Human Factors, 20*(5), 603-609.

Overview:

This paper investigates the effect of load stress and speed stress on visual tasks where attention must be shared across several channels. *Load stress* is stress caused by increasing the number of channels over which information is presented (Gawron, 2008), and *speed stress* is the stress caused by changing the rate of signal presentation (Sanders & McCormick, 1993). Goldstein and Dorfman suggested that the information processing requirements change with changes in speed and load stress.

This effect was studied using an experiment where subjects were required to respond to dynamic visual stimuli which entered critical zone in each of three visual displays. Various combinations of speed and load stress were presented in a time sharing task where subjects were required to respond quickly to frequent, non-predictive signals. Load stress was increased by altering the number of displays that the subjects had to interact with, and speed stress was controlled by altering the rate of signal presentation. It was found that both types of stresses contributed to performance. An increase of load stress and/or speed stress led to an decrease in performance.

From these results, Goldstein and Dorfman concluded that performance was most negatively affected by load stress, particularly in conditions of combined high load stress and high speed stress. However, they suggest that practice and predictive cueing can help to alleviate the effect of the high load condition. Lastly, the authors warn that speed stress and load stress should not be considered independent of each other when repeating similar experiments.

Conclusions:

This paper suggests that interface designers should work towards reducing load stress and speed stress in order to maximize operator performance. Thus, the number of channels where information is presented should be reduced, and the frequency of presenting information should be reduced.

Reference:

Healey, C. G., Booth, K. S., & Enns, J. T. (1996). High-speed visual estimation using preattentive processing. *ACM Transactions on Computer-Human Interaction*, 3(2), 107-135.

Overview:

This study demonstrated a new form for performing rapid numerical estimation through pre-attentive processing. The authors defined pre-attentive processing as " an initial organization of the visual field based on cognitive operations believed to be rapid, automatic, and spatially parallel" (e.g. hue, orientation, size, motion and intensity). The authors hypothesized that pre-attentive vision can result in rapid and accurate visual analysis in visual displays. In the context of numerical estimation, this study examines two pre-attentive features which are hue and orientation to determine whether pre-attentive estimation is possible or not. Experiments involved subjects interpreting salmon migration simulations presented on visual displays in terms of percentage values. Results indicated that "rapid and accurate" estimations were possible using hue and orientation pre-attentive features. In addition to these results, the authors provided a chart summarizing various researchers that used the following visual features to perform pre-attentive tasks.

Table A-5: List of various researchers that used the following visual features to perform pre-attentive tasks

Feature	Author
line (blob) orientation	Julész & Bergen [1983]; Wolfe [1992]
length	Triesman & Gormican [1988]

width	Julész [1985]
size	Triesman & Gelade [1980]
curvature	Triesman & Gormican [1988]
number	Julész [1985]; Trick & Pylyshyn [1994]
terminators	Julész & Bergen [1983]
intersection	Julész & Bergen [1983]
closure	Enns [1986]; Triesman & Souther [1985]
colour [hue]	Triesman & Gormican [1988]; Nagy & Sanchez [1990]; D'Zmura [1991]
intensity	Beck et al. [1983]; Triesman & Gormican [1988]
flicker	Julész [1971]
direction of motion	Nakayama & Silverman [1986]; Driver & McLeod [1992]
binocular luster	Wolfe & Franzel [1988]
stereoscopic depth	Nakayama & Silverman [1986]
3-D depth cues	Enns [1990]
lighting direction	Enns [1990]

Conclusions:

The natural pre-attentive processing capabilities within humans should be taken advantage when designing visual displays to ensure that the operator's attentional resources are being allocated optimally.

Reference:

Helbig, H. B., & Ernst, M. O. (2007). Knowledge about a common source can promote visual – haptic integration. *Perception*, *36*(10), 1523-1533.

Overview:

This paper addresses past research which suggests that when two signals come from the same object, integration is supported even if the signals are in spatial conflict. The purpose of this is to resolve conflicting opinions that multiple signals from the same object can promote sensory integration.

Three experiments were completed to evaluate this issue. In all three experiments, subjects were required to respond to the shape of an object by selecting a comparison object which matched in shape. However, for determining the interaction between tactile and visual stimuli, there was a conflict introduced between the visual and tactile properties of the object. The first experiment consisted of two conditions. In the first, subjects had a direct view of the object touched. In the second condition, mirrors were utilized which created a spatial separation between the viewed and felt object. This experiment was designed to test whether previous awareness that the two sensory signals arose from the same object supports integration, despite the fact that the two signals are presented at conflicting locations. In this experiment, subjects were required to report the perceived shape. For the second experiment, the authors suggested that perhaps the determination of the shape property promoted sensory integration. Thus, for the second experiment, subjects were asked to complete the same task as experiment one, but report on the visual of haptic shape percept instead. The third experiment was presented as a control study, which verified that in the absence of secondary knowledge about a common source, sensory integration breaks down when the multimodal signals are in spatial conflict.

From the three experiments, Helbig and Ernst found the existence of a mutual biasing effect of shape information from the visual and tactile modalities. These findings were not dependent on the presence of either of two cue conditions. These findings suggest that previous knowledge regarding the object properties can help to promote integration of sensory modalities, despite the presence of spatial discrepancies between the visual and tactile modalities.

Conclusions:

The work completed by Helbig and Ernst supports the existence of sensory bias, where the bias is a function of the property being determined. Understanding bias situations can help interface designers to determine which modalities should be used to present certain properties of a stimulus or object.

Reference:

Ho, C., Santangelo, V., & Spence, C. (2009). Multisensory warning signals: When spatial correspondence matters. *Experimental Brain Research*, 195(2), 261-272.

Overview:

The goal of this paper is to show the effectiveness of unimodal and bimodal audiotactile stimuli in luring the subject's spatial attention away from a highly perceptually demanding central rapid serial visual presentation (RSVP) task. The unimodal and bimodal audiotactile stimuli were not relevant to the task being completed by the subject.

Three experiments were completed, two of which are relevant and explained below. In the first, subjects were asked to provide speeded elevation discrimination responses to peripheral visual targets, where the targets were preceded by auditory stimuli. These stimuli were either presented alone or were combined with centrally presented tactile stimuli. The purpose of this experiment was to study the role of spatial separation in multisensory audiotactile interactions. Specifically, the goal was to compare the relative effectiveness of unimodal and bimodal audiotactile stimuli in two conditions: no-load and high perceptual load. In the second experiment, the spatial auditory stimuli were either presented alone or in a combination with a tactile stimulus originating from the same spatial location. The purpose of this experiment was to investigate whether audiotactile directional congruency was effective in enhancing the subject performance.

The results from the first experiment indicated that the unimodal auditory stimuli were only effective when subjects were not involved in the RSVP task. In addition, the bimodal audiotactile stimuli did not show any performance change through the different conditions. These findings contrasted with the thought that audiotactile cues may increase performance in higher perceptual loading tasks. Ho, Santangelo and Spence therefore suggested that the audiotactile integration of cues may require that the auditory and tactile components of the cues originate from the same spatial direction.

The results from the second experiment differed from those in the first experiment, because the bimodal audiotactile stimuli were effective in capturing the subjects' spatial attention from the concurrent RSVP task. These results further supported the claim that auditory and tactile stimuli should be presented from the same direction.

Conclusions:

Ho, Santangelo and Spence suggest that interface designers need to consider the spatial arrangement of multisensory information in their designs. Specifically, tactile information presented on the body surface may not be effective if its spatial directionality is shifted relative to its auditory pairing.

Reference:

Ho, C., & Spence, C. (2009). Using peripersonal warning signals to orient a driver's gaze. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 51(4), 539-556.

Overview:

This paper addresses recent findings that have shown that the human brain considers stimuli occurring in the *peripersonal space* as more relevant and attention-demanding. Ho, Spence and Kingdom investigated this concept with a focus on the application of designing warning signals. Three experiments were completed which assessed the speed (reaction time) at which participants could initiate head-orienting responses following the occurrence of spatial warning signals.

The goal of the first experiment was to determine the relative speed at which subjects could initiate speeded head-oriented responses (to the left or right), starting from a facing forward position. This experiment tested the effectiveness of various unimodal warning signals in causing a head movement response in the direction of the danger requiring attention. The goal of the second experiment was to evaluate the relative effectiveness of various unimodal signals (auditory, visual, and tactile) in alerting and capturing a driver's attention in the appropriate direction. The third experiment evaluated the relative effectiveness of various warning signals in redirecting a subject's gaze back to the central tasks while the subject was involved in a secondary task.

Results indicated that subjects began their head turning movements and made speeded discrimination or braking responses significantly faster following the simulation of a nearby rear auditory warning signal than following the display of either a far frontal auditory warning or a vibrotactile warning signal presented at the waist or a peripheral warning signal (signals not directly in front of the participant).

Conclusions:

Ho, Spence and Kingdom suggest that multimodal warning systems designed around the constraints of the human brain provide a greater potential for information communication. Their results support earlier work that warning signals which activate the brain's defensive circuit for self-protection (peripersonal space) offer an effective means of alerting operators of errors.

Reference:

Johnson, J. S., Woodman, G. F., Braun, E., & Luck, S. J. (2007). Implicit memory influences the allocation of attention in visual cortex. Psychonomic Bulletin & Review, 14(5), 834-839.

Overview:

The authors hypothesize that implicit memory influences the allocation of attention for contextual information. In order to test this hypothesis, the authors required subjects to search for a rotated "T" target amongst rotated "L" distracters. Subjects were told to respond to the "T" target by pressing one of two buttons, indicating whether the target pointed to the left or right. This search task was said to be known for requiring spatial attention. An instrument called N2pc component a "well-validated electrophysiological signature of focusing attention" was utilized to observe shifts of attention. Since it is an established finding that shifts of covert attention follow eye movements, N2pc effects are a reliable source that is capable of reflecting shifts of attention directly to the target. It was observed that reaction times were significantly faster for targets appearing in repeated arrays than for novel arrays.

The authors state that "the use of implicit memory to control attention may play a key role in real-time sensorimotor processing because it obviates the need to use prefrontal executive systems to guide an explicit memory search process, making perceptual processing faster and freeing executive systems to focus on other tasks. This idea (that implicit memory can be used to process information without using other memory sources) complements previous research indicating that attention can be focused on objects to discriminate them without storing them in visual working memory."

Conclusions:

These findings show potential for using the advantages of implicit memories to increase performance while requiring little attentional resources. For example, in terms of the current project's objectives, interface designers can present information in arrays that the user is familiar with so that he/she can use their implicit memory as a source to interpret the information.

Reference:

Kitagawa, N., Zampini, M., & Spence, C. (2005). Audiotactile interactions in near and far space. *Experimental Brain Research*, 166, 528-537.

Overview:

This paper presents the results of an experiment which studied the audio-tactile spatial interactions in the region behind the head. Two experiments were completed for the investigation. In the first, the subjects were required to make unspeeded temporal order judgments (TOJs) of pairs of auditory and tactile stimuli. These stimuli were presented at varied stimulus onset asynchronies. In the second experiment, auditory stimuli were introduced to the discrimination task to distract the subjects. This auditory interference was created using two large speakers located to the left and right behind the subjects. The purpose of this task was to show that speeded discrimination responses (localization of the stimulus to either the left or right of the body) to electrocutaneous targets (a type of electrical stimulus placed on the skin which differ from vibrotactile stimuli which are vibrating stimuli placed on the skin) were also changed by the spatial congruency of auditory distracters presented behind the head.

From the first experiment, it was seen that subjects provided more accurate responses when the stimuli were presented from different sides of the head than from the same side. However, the second experiment showed that when auditory interferences were presented on the opposite side of the electrocutaneous target, response times increased and accuracy decreased in the localization task when compared to congruent (same side) presentations of auditory and electrocutaneous stimuli. This negative effect became stronger when white noise distracters were presented close to the head (20cm), than when they were presented further from the head (70cm). On the contrary, pure tone inferences showed a smaller effect to the distraction, and showed no change as a function of distance from the head.

The findings by Kitagawa, Zampini and Spence conducted research on how different types of auditory stimuli can affect information processing. The results indicated that white noise stimuli presented in the vicinity of the back of the head affected tactile response times and accuracy more strongly than white noise presented far from the head. The effect was also strong as compared with pure tone stimuli, regardless of the distance of the stimulus from the head. The collection of these finding shows that audiotactile interactions in information processing are stronger for complex sounds, such as white noise. Also, these interactions are strongest when the stimulus is presented behind the head in the peripersonal space.

In addition, from the first experiment, it was shown that subjects were more accurate when the stimuli were presented from a variety of spatial positions, rather than when the stimuli were presented in the same position behind their heads. This suggests that audiotactile interactions occur at a preattentive perceptual level, instead of solely at a decisional level. In the second experiment, it was found that audiotactile interactions for stimuli placed behind the head also affected performance in a speeded spatial discrimination task. It was found that electrocutaneous (left versus right earlobe) discrimination performance was worsened in situations where the auditory interferences were presented on the opposite side of the target, compared to situations where the interference was presented on the same side. The authors point out that this provides further evidence for auditory-tactile interactions, which replicates findings that had been found in earlier studies. However, the authors did not describe any particular reasons why this interaction may occur.

<u>Conclusions:</u>

The study by Kitagawa, Zampini and Spence shows support for the above mentioned fact that signals placed in the peripersonal space can be more effective. However, in this study, it was shown that distracters in the peripersonal space can significantly affect response time and accuracy. Thus, when designing multimodal interfaces, it is important to place warning signals in the peripersonal space, but important to prevent distracter signals from occurring in the same space.

For example, Gilliland and Schlegel (1994) presented several studies which investigated the effectiveness of head-mounted tactile devices, used for presenting localizable signals to pilots by vibrating difference positions of the head. However, pilots usually receive extensive auditory information using headphones or radios. The findings by Kitagawa, Zampini and Spence suggest that there may be a potential conflict between the information showed over these two channels.

The authors emphasize the concern that interface designers should be aware of multisensory constraints on information processing.

Lastly, this research suggests that the characteristics of auditory stimuli (e.g. white noise versus pure tone) may affect the perception and subsequent information processing of the event. This information is important for interface designers because it provides them with information on how to select the most effective auditory stimulus for its applications.

Reference:

Koppen, C. M., & Spence, C. (2007c). Spatial coincidence modulates the Colavita visual dominance effect. *Neuroscience Letters*, 417(2), 107-111.

Overview:

After Colavita coined the “Colavita Visual Dominance Effect,” many researchers revisited this finding in attempts to examine this phenomenon in depth. Koppen and Spence have demonstrated that there are various factors that modulate the magnitude of the Colavita visual dominance effect. In this paper the authors propose that spatial coincidence modulates the Colavita visual dominance effect. Spatial coincidence is defined as something occurring in the same spatial location. Auditory, visual, and bimodal stimuli were presented to participants and the experiment required them to respond to the stimulus by either pressing an "auditory response key" or a "visual response key" or both keys. In regards to bimodal trials, participants responded more often to the visual stimulus, exemplifying the Colavita visual dominance effect. However, when the auditory and visual components of the bimodal targets were presented in different spatial locations (13° or 26°), the Colavita visual dominance effect was significantly less apparent. Koppen and Spence concluded that spatial coincidence modulates the Colavita visual dominance effect. A possible explanation for this modulation that Koppen and Spence pointed out is that research has demonstrated that visual performance in terms of response latencies is poorer in the periphery compared to central vision.

Conclusions:

This suggests that interface designers should convey visual displays/information near each other because of possible slow response latencies. This is also another important aspect to consider when interface designers are using the visual modality as a channel to present information because the user’s ability to comprehend/respond to information can be affected by spatial presentation.

Reference:

Koppen, C., & Spence, C. (2007a). Assessing the role of stimulus probability on the Colavita visual dominance effect. *Neuroscience Letters*, 418(3), 266-271.

Overview:

This paper claims that stimulus probability modulates the Colavita visual dominance effect. In previous Colavita visual dominance studies, the proportion of audio, visual and audiovisual targets were 40A:40V:20AV however this study manipulated the proportion of targets to 25A:25V:50AV which resulted in the magnitude of the Colavita effect to significantly decrease. This finding appears to be consistent with literature on attention stating that an increase in the frequency of specific targets (e.g. bimodal targets), will direct the participants' endogenous attention towards that specific target which will improve performance in speeded discrimination response tasks. Thus, the authors concluded that by increasing the probability of bimodal stimuli, the Colavita effect is reduced.

Conclusions:

The authors have demonstrated that by manipulating the probability of stimulus can affect the expectancies of participants. Thus, interface designers should keep in mind that users will respond more often or have shorter response latencies to expected stimuli. When designing an interface, the user should be familiar with various information the interface may present in different situations so that he/she is aware of different things that can occur and possible probabilities of various occurrences. This can help the user interact with the interface more effectively, thus improving overall performance.

Reference:

Koppen, C., & Spence, C. (2007b). Audiovisual asynchrony modulates the Colavita visual dominance effect. *Brain Research*, 1186(1), 224-232.

Overview:

This paper claims that audiovisual asynchrony modulates the Colavita visual dominance effect. Similar to the other papers Koppen and Spence published regarding the Colavita effect ("Assessing the role of stimulus probability on the Colavita visual dominance effect" and "Spatial coincidence modulates the Colavita visual dominance effect") auditory, visual, and bimodal audiovisual stimuli were presented to participants. They had to respond to the stimulus with the appropriate key (either a visual response key or auditory response key or both). It was observed that participants responded to the visual component of the bimodal targets more often than the auditory component. When the stimulus onset asynchrony (*SOA*) between the visual and auditory component of the bimodal targets varied, the Colavita effect began to disappear as participants reliably reported the auditory component appearing first. Thus, the authors concluded that these results exemplified the modulation of the Colavita visual dominance effect caused by the temporal order of the audiovisual bimodal targets.

Conclusions:

Through the various experiments Koppen and Spence have conducted, if we desire to present information to the operator through a modality other than vision (see Occelli, O'Brien, Spence, & Zampini, 2010 for an example of the visuotactile Colavita effect), we should manipulate the circumstances so that the visual dominance tendency is reduced. We can use these studies as exemplars of manipulations/circumstances that the visual dominance effect was attenuated. For instance, if an interface designer wanted to use the visual dominance phenomenon to his/her advantage, he/she could present information through the visual modality but must ensure that the modulating factors Koppen and Spence examined are not present (e.g. spatial coincidence, probability, audiovisual synchrony). By taking these factors into consideration, interface designers could prevent the visual dominance effect from being attenuated.

Reference:

Lederman, S. J., Thorne, G., & Jones, B. (1986). Perception of texture by vision and touch: Multidimensionality and intersensory integration. *Journal of Experimental Psychology: Human Perception and Performance*, 12(2), 169-180.

Overview:

This paper addresses the differences in the ways that the visual and tactile modalities utilize textural information. A series of six experiments are presented in this study.

In experiments one and four, subjects were asked to determine an undetected texture discrepancy (where the visual texture and the tactile texture differ, but is not detectable by observers) between the visual and tactile modalities in terms of the spatial density of the pattern elements. In the second and third experiments, the subjects were asked to resolve the same discrepancy as in experiment one, but also were asked to determine the roughness of the surfaces. In experiments five and six, average spatial density and average roughness were evaluated, respectively, of pairs of textured surfaces, one presented to the visual modality and the other presented to the tactile modality.

These experiments showed that although both the tactile and visual modalities are used to determine surface texture, the relative weighting that observers apply to these two modalities changes depending on the texture dimension being analysed. These results challenged the previous suggestion of visual dominance and showed that both touch and vision contribute to the perceived spatial density and roughness of raised surface patterns. Lederman, Thorne and Jones suggested that it is vital to compare the differences in processing strategies caused by the multidimensional nature of texture perception. Lastly, it was demonstrated that the processing of spatial density and roughness information by the visual and tactile modalities may be described using a weighted average model.

Conclusions:

This paper shows how the tactile and visual modalities can be combined to optimize the detection of different parameters of a textured surface. This information is useful for the design of multimodal interfaces where texture information needs to be relayed to the operator.

Reference:

McDonald, J. J., Teder-sälejärvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature*, 407, 906-908.

Overview:

This paper presents psychophysical evidence which shows that an abrupt auditory stimulus improves the detection of a subsequent visual stimulus. In this paper, both stimuli were designed to originate from the same spatial location. McDonald, Teder-sälejärvi and Hillyard addressed the question of whether a stimulus in one sensory modality automatically attracts attention to another stimulus in a different modality, which occurs in the same spatial location. The purpose of this work was to see if the perception of a spatial event can be enhanced using multiple modalities.

The researchers used signal detection measures, instead of reaction times, investigate the subjects' effectiveness in completing the task. This allowed for the results to show a separation of perceptual and decision-level effects of attention. In signal detection theory, the d' parameter indicates the ability of the subject to distinguish a sensory event from its background. In the context of this study, the d' parameter should be larger for the flashes that occur close to the previous sound, if the involuntary orientation of attention to the location of an auditory stimulus is supported by the visual perceptual processes.

Two cross-modal cueing experiments were completed. The first experiment, a nonpredictive spatial auditory cue was provided at an offset from the fixation point. This event was followed by a visual mask at either the same location (valid test) or at a different location (invalid test). The second experiment was similar to the first, but in this case the response accuracy measured over the speed.

The study found that a sudden auditory stimulus improves the detection of a flash following the auditory event, at the same spatial location. The researchers also mentioned that previous research has found that an irrelevant auditory stimulus can modify the perception of concurrent or subsequent visual stimuli (e.g. an increase in intensity of a flash). In this experiment, the authors found evidence that this effect occurred both when the auditory and visual stimuli occur at the same location, and when the locations are different. However, this effect occurred only when the subjects were focused on the visual stimuli. The authors suggest that the effects of an auditory stimulus on the processing of concurrent and subsequent visual stimuli are caused by separate neural mechanisms.

Conclusions:

This article indicates how an auditory stimulus can improve the detection of a visual stimulus. In the design of multimodal interfaces, this concept can be used to enhance a visual warning signal or cue.

Reference:

Ma, W. J., & Pouget, A. (2008). Linking neurons to behaviour in multisensory perception: a computational review. *Brain research*, *1242*, 4-12. Elsevier B.V.

Overview:

This paper presents a review of past work in Bayesian modeling for information integration across multiple senses. Past research comprising of psychophysical and physiological findings have focused on two major areas: *how* do humans integrate information, and *when* do humans integrate information?

First, the paper presents the optimal cue integration Bayesian model, also referred to as the maximum-likelihood estimation. This method assumes that a common source present. Then, a small conflict, which cannot violate the common-source assumption, is introduced into the system. From this, an estimate of the stimulus is determined from both cues, it is determined from the thinking that the percept will lie somewhere between the percepts determined from each cue individually. It is assumed that the higher weight will be given to the most reliable cue.

The review also explains the causal inference model, currently considered to be the best model for predicting multisensory interactions. For this model, the observer considers tow possible hypotheses: multisensory signals have a common cause, or multisensory signals have separate (two) independent causes. For each event, the observer determines the probability for each hypothesis and uses information regarding system noise and prior knowledge to reach a decision.

Ma and Pouget also suggest that there is much more work to be completed in the area of Bayesian modeling, such as determining models for more complex stimuli and for conflict situations.

Conclusions:

Bayesian modeling can assist interface designers by allowing them to predict the performance of multisensory cues without using an experimental set-up. Also Bayesian modeling is not a substitute for experiments; they can help predict performance during the design process.

Reference:

Moray, N. The role of attention in the detection of errors and the diagnosis of failures in man-machine systems. *Rasmussen, J., Rouse, W.B. (1981). Human Detection and Diagnosis of System Failures. Proceedings of a NATO Symposium (Pp.185-198). New York, NY: Plenum. x+716pp.; Human Detection and Diagnosis of System Failures. Proceedings of a NATO Symposium, 4-8 Aug. 1980, Roskilde, Denmark. NATO; Riso Nat. Lab. Denmark*, 185-198.

Overview:

Over twenty years ago, very little research existed which addressed the effect of attention on error detection and diagnosis. Of course, much work has been completed since then in this area, but Moray presented several fundamental theories which present-day research works from. In his research, Moray investigated the need to pay attention to several sources of information, as well as to the details of the information received from those sources, specifically for the application of operator issues when controlled automatic and semi-automatic systems.

At this time, multimodal interfaces were not a commonality, so visual displays were the focus of Moray's research. Moray addressed one attentional limit as the rate at which sources of information can be sampled by an operator (referred to a speed stress in other parts of this report). He reported that eye movements cannot be made at a rate significantly faster than two fixations per second. Also, it was noted that focused attention is required for accurate pattern recognition, even if the rates of change can be detected by the periphery of vision. Moray also references past studies and indicates that operators with long experience of a system can develop unconscious scanning patterns which are nearly optimal.

For interface design, Moray indicates that only two samples per second can be taken using the visual modality, which translates a state variable bandwidth requirement of 0.01Hz for the sampling to be adequate. Notably, in many highly dynamic situations, such as the final stages of landing an aircraft, the bandwidth far exceeds this preferred value. Also, as the required sampling frequency of a source increases, so does the likelihood that the attention will be overloaded. Naturally, the more loaded the attention is, the less likely that an observation of an abnormal variable will occur. In this paper, the use of multisensory stimuli is suggested by adding an additional channel of auditory stimuli. Previous work suggested that the maximum of four auditory signals should be used to prevent overload situations.

With regards to interface design, Moray indicates that high speed and high accuracy cannot be attained simultaneously. Rather, there is a speed-accuracy trade-off function for the observation of dynamic functions. The author suggests that large complex systems can be composed of a number of subsystems, which contain variables which may or may not be connected to each other. Despite this, it is possible that humans can create correlations between the variables, and thus create optimal strategies for examining the system as a whole. Moray's fundamental inclination is that following an abnormal observation, highly correlated sources should be sampled. Although this may be advantageous in some situations, it is possible that "cognitive tunnel vision" will result.

From his hypotheses and research, Moray determines a set of design criteria for man-machine systems:

- Minimize the number of displays
- Let the system monitor the operator
- Minimize data acquisition time
- Use predictor displays
- Make the system demand interrogation during diagnosis

Conclusions:

Although more recent work should be considered when designing multimodal interfaces, Moray's work and proposed design constraints provide a basis for how man-machine systems should be designed to reduce failures.

Reference:

Moray, N., & Inagaki, T. (2000). Attention and complacency. *Theoretical Issues in Ergonomics Science*, 1(4), 354-365.

Overview:

This paper addresses the issue of complacency in monitoring tasks. The definition of complacency is: self-satisfaction which may result in non-vigilance based on an unjustified assumption of satisfactory system state. Moray and Inagaki address the concerns from previous research which suggest that complacency cannot be proved unless optimal (best possible performance for a human observer) behaviour is specified as a benchmark. An experimental evaluation is completed in this paper which shows that even when operators made use of optimal scanning and monitoring techniques, not all signals can be detected. Thus, it is safe to assume that there will be times when an operator would miss a target of interest.

The results from three experiments showed that there are situations where optimal monitoring will cause critical signals to be missed. The only way to guarantee that all signals are detected is to devote the attention entirely and continuously to one process, where the critical signals are expected to appear.

Moray and Inagaki proposed the following question: at what frequency should a 100% reliable source be sampled? One approach would be to model the source, both causally and mathematically. The model would account for a worst case situation where the operator would be required to intervene and take action towards preventing the fault from becoming a disaster. The authors proposed a model of how frequently a fault-less system (i.e. 100% reliable) should be sampled (looked at),

$$f \cong (T - t)w$$

where T is the time from the occurrence of a fault until the dangerous consequences are unavoidable (the incident is unrecoverable), t is the time required to take action to prevent the unrecoverable consequences, and w is a weight related to the severity of the consequences of an unrecoverable accident. Since many multimodal interfaces increase the number of potential sources of information (because information can now be presented in different modalities), it is important to determine how frequently an operator may sample one of these information sources. This can assist with the measurement of complacency, as mentioned above, and it can also be used to determine the perceptual workload of the operator.

Also, the authors noted that to claim complacent behaviour is the same as blaming the operator for failing to detect signals. However, they also claimed that the existence of complacent behaviour is inherently caused by poor system design. Thus, interface designers must take more care to design systems where complacency is less likely to occur or where there is redundancy for missed targets, since it is inevitable that operators will miss some targets even if optimal scanning behaviour is maintained. The methods for accomplishing this still need to be investigated. Also, highly reliable sources should be replaced by warnings, since it is highly probable that they will not be monitored.

Conclusions:

Moray and Inagaki suggest that since optimal behaviour cannot guarantee the timely detection of all signals, operators should not be expected to detect all faults. This concept is applicable to the design of multimodal interfaces, because it indicates that not all information should be presented as monitoring tasks. For example, it would be more beneficial to present a highly reliable source using warning signals as opposed to presenting a signal that required continuous checking. The authors also suggest that the only possibility for directing attention effectively in all situation is be the use of "attentional interrupts" which override any existing strategy when the critical signals occurs. However, these attentional interrupts must not provide false alarms (possibly by increasing the reliability of the alarm), or they will not be trusted. If an operator no longer trusts the system, then they may disregard any further information that comes from the unreliable source.

Reference:

Pattyn, N., Neyt, X., Henderickx, D., & Soetens, E. (2008). Psychophysiological investigation of vigilance decrement: Boredom or cognitive fatigue?. *Physiology & Behavior*, 93(1-2), 369-378.

Overview:

This paper addresses human-related issues during tedious monitoring tasks. The goal of the paper was to address three research questions. First, which type of attention is more susceptible to vigilance decrements due to the amount of time spent on the task, endogenous or exogenous? Second, can measures of autonomic arousal address the issue of decreased workload leading to the inability to sustain mental effort? Lastly, do the measures show a different effect for endogenous versus exogenous attention?

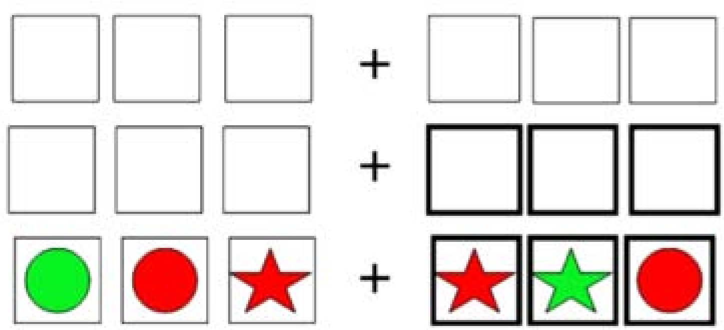


Figure A- 39: Course of a valid trial in the exogenous condition. Upper row: fixation display; middle row: exogenous valid cue display (800 ms); the boxes are brightened in the location where the target will appear in the following display; bottom row: target display, with the green star being the target.

An experiment was conducted in an attempt to answer these research questions, where subjects were required to respond to a concurrently cued search task. For each trial, subjects were exposed to three types of displays: fixation, cue, and target display (See Figure A- 39).

In the exogenous condition, either the three right boxes or three left boxes would increase in brightness, which would suggest where the target would likely appear. In the endogenous condition, the plus signs were replaced with arrows, which would suggest the possible location of the future target. In both conditions, it was possible for the cue to be valid or invalid. The researchers analyzed cardio-respiratory signals, response times, and subjective data to reach their conclusions. The subjective data included questions about whether a disturbing factor was experienced during the experiment, how subjects had managed with the long time-on-task, and whether a strategy was implemented for concentration. Also, ratings were collected which evaluated the subjects' thoughts on their performance.

Pattyn, Neyt, Henderickx and Soetens found that endogenous and exogenous attention showed a different evolution over time. The response time data indicated that there was a dual effect from time-on-task. Each participant was tested over a period of 1.5 hours which was divided into three blocks of 30 minutes. There was a slowing of response from the first time block to the second time block, which occurred for both the endogenous and exogenous conditions. However, this effect was stronger for the endogenous condition. Also, a larger validity effect ("the difference between RTs after invalid cues and RTs after valid cues" (p. 373)) over time occurred for the endogenous condition, which is strongly due to slower response times after invalid cues, which were presented in the second and third time block. This result indicates that there may be a higher cost associated with changing the focus of attention from one side of the fixation to the other side.

There was a general slowing in response times after the first time block. According to past literature, the time until response times start to slow is approximately twenty to thirty minutes. Furthermore, an additional switching cost occurred after an invalid cue was presented in the endogenous condition. These results show that performance is more efficient following exogenous cueing. This increase in performance is characterized by faster response times, smaller error rates, and less vulnerability to time-on-task.

Also, the subjective responses from the participants in evaluating their own performance was best at the first time block, and worsened after the second time block. Participants communicated a feeling of being bored, and did not feel as though they were under a high mental effort. There was no physiological difference found between the endogenous and exogenous conditions.

Conclusions:

This study investigates the issues surrounding underloading and loss of attention in tedious monitoring tasks. These issues are strong problems for interface designers, as human boredom can lead to an increase in human error. This research helps us to understand situations where lack of attention occurs.

Reference:

Roach, N. W., Heron, J., & McGraw, P. V. (2006). Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. In *Proceedings of the royal society B* (Vol. 273, pp. 2159-68). The Royal Society.

Overview: This paper investigates interactions between auditory and visual rate perception. In the study presented, subjects were asked to provide responses in one modality while ignoring contradictory information presented in another modality. From this study, the authors determine a new Bayesian model which addresses issues with the precursor model, the maximum-likelihood estimation. From investigation on auditory and visual rate perception, Roach, Heron and McGraw found that a gradual transition between partial cue integration and complete cue segregation with intermodal discrepancy existed. These results were not in accordance with the maximum-likelihood estimation model. In an attempt to explain these results, the authors proposed a new Bayesian model, which considers prior knowledge about the connection between auditory and visual rate signals. Thus, a strategy is derived which balances the benefits accumulated from integrating sensory estimates from a common source against the cost of similar information related to independent objects or events.
Conclusions: More recent research indicates that the causal inference model fits experimental data than other Bayesian models. However, an understanding of past work in Bayesian modeling is important in deriving improved models for improving the predictability of multisensory interfaces.

Reference: Santangelo, V., Fagioli, S., & Macaluso, E. (2010). The costs of monitoring simultaneously two sensory modalities decrease when dividing attention in space. *NeuroImage*, 49(3), 2717-2727.
Overview: This paper addresses and challenges the concept of stimulating or attending to different senses at one single location. In the past research, this concept has been reported to be advantageous, but Santangelo, Fagioli and Macaluso suggest that the in-parallel processing of two sensory modalities can be more effective when a person's attention is spatially divided, instead of focused. This suggestion was experimentally tested, where subjects were asked to monitor visual and auditory stimuli concurrently at either one location in two opposite hemifields, or one modality at one or two locations. These options corresponded to focused attention, divided attention, and mixed attention, respectively. The behavioural results indicated that the division of attention across space resulted in smaller

costs of monitoring two modalities over one modality. Also, fMRI results indicated that brain activity in the dorsal fronto-parietal regions increased both for attending to multiple locations and for monitoring multiple modalities, which suggests that a common system is used for processing an increasing number of attended streams. Also, neuroimaging data showed that there was an increased activity in the posterior-parietal cortex for the divided attention condition, but no specific region was used in the focused attention condition. These two findings support the theory of supramodal control for multisensory processing. To account for the above results, the authors suggest that supramodal control and the integration of spatial information impede the selection of individual sensory streams in the focused attention condition, and the utilization of modality-specific resources and the engagement of the posterior-parietal cortex allows in-parallel processing in the divided attention condition.

<u>Conclusions:</u>

The results found by Santangelo, Fagioli and Macaluso suggest that multisensory cues can be effective for cases where an operator's attention is spatially divided. This is applicable to multimodal interface design, because the use of this concept would allow for operators to monitor two channels of information at different locations concurrently and effectively.

<u>Reference:</u>

Scott, J. J., & Gray, R. (2008). A comparison of tactile, visual, and auditory warnings for rear-end collision prevention in simulated driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(2), 264-275.

<u>Overview:</u>

The purpose of this paper is to investigate the effectiveness of multimodality warnings for rear-end collisions, as a function of warning timing in a driving simulator. Since the use of in-vehicle information and entertainment systems can lead to driver inattention, it was pertinent to examine which types of warnings are more effective at capturing the driver's attention in case of an emergency (e.g. faster response time). Subjects were placed in a fixed-base driving simulator and were instructed to follow a red lead car on a rural two-lane road. They were directed to drive in their own lane and not pass the lead car. The drivers were presented with counterbalanced blocks of visual, auditory, and tactile warning, plus a no-warning, baseline condition. The warnings were activated when the time-to-collision (TTC) reached a critical threshold of three to five seconds. The response time of the driver was captured from the time that a warning was initiated below the critical threshold until brake initiation. For the purpose of simulating real-world driving scenarios, drivers listened to background music of their preference at 60dB to engage the auditory system, and occasional opposing roadway traffic was presented to engage the visual system.

Scott and Gray found that the response time of the driver was the lowest using a tactile warning, and the highest in the no-warning condition. Also, the tactile response times were significantly shorter than for the visual modality. This suggests that tactile warning signals provide faster response times than visual warnings. There was not a significant difference between the auditory and tactile warnings, which may be a result of the auditory loading.

Also, there was a statistically significant effect in the response times between the three second TTC warning and the five second TTC warning. The response times were shorter for the three second TTC condition. Scott and Gray suggest that this may be due to the fact that the drivers had more time to make decisions and thus often opted to coast before applying the brakes.

The findings of this study show that tactile stimuli for warning application reduced driver responses times when compared to visual or auditory stimuli.

Conclusions:

This study suggests that the tactile modality produces the fastest response times to alerts when compared to visual or auditory stimuli, especially when the operator's attention is directed elsewhere.

Reference:

Spain, R. D., & Bliss, J. P. (2008). The effect of sonification display pulse rate and reliability on operator trust and perceived workload during a simulated patient monitoring task. *Ergonomics*, 51(9), 1320-1337.

Overview:

Sonifications were said to be a useful tool to promote "eyes-free continuous monitoring without disrupting attentional focus" thus it is a useful pre-attentive processing tool. Research findings have demonstrated that manipulations of pulse rate can portray urgency levels effectively. Since the rate of information presentation can affect the operator's performance in terms of the operator's ability to process information, it is important for interface designers to know optimal levels of information presentation to the operator. Thus, this paper explores the influence of sonification signalling rate and system reliability effect the operator's mental workload and trust in sonification. The experiment conducted consisted of three different sonification pulse rates and two different levels of system reliabilities which participants were assigned to randomly. The three levels of sonification pulse rates included 40 pulses per minute (ppm); 60 ppm; and 80 ppm. The two system reliability levels included 40% true alarms and 60% true alarms (e.g. in the case of 40% true alarms, 4 out of 10 alarms would represent a true problem with the patient's blood pressure status and likewise with the 60% true alarm condition). Participants were required to monitor the status of a patient (secondary task) while attending to a primary task. The status of a patient was presented in the form of various frequency auditory pulses to represent the patient's blood pressure. An increase in the rate of the auditory pulse indicated a potential problem with the patient's blood pressure. Participants were told they should attend to patients when felt necessary (in accordance to auditory pulse rate) in the various conditions explained earlier. Overall results demonstrated that participants displayed greater trust when they encountered the more reliable systems (60% true alarm condition) compared to the less reliable system (40% true alarm condition). In addition, participants also displayed greater trust when they encountered the 670 ppm condition compared to the 40 ppm condition. They also exhibited less perceived amount of workload n the 60 ppm condition compared to the 40 ppm and 80 ppm condition. A possible explanation for this result is that participants may have perceived the 80 ppm condition has being "inundated " with too much information and the 40 ppm condition as not getting enough information fast enough, thus the 60 ppm condition may be a comfortable signalling rate. Another possible explanation the authors pointed out the 40 and 80 ppm conditions placed a "greater burden on the working memory" than the 60 ppm condition. Overall, this paper and previous research has demonstrated how pre-attentive resources can allow operators to monitor various levels/statuses without tapping into attentional resources.

Conclusions:

Pulse rates manipulations can be used to portray urgency levels in auditory/sonification displays. The optimal level of pulse rate in the medical context field in terms of mental workload, trust and interpretability appears to be 60 ppm.

Reference:

Spence, C., & Driver, J. (1996). Audiovisual links in endogenous covert spatial attention. *Journal of experimental psychology. Human perception and performance*, *22*(4), 1005-30. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8756965.

Overview:

This paper presents an investigation on the existence of cross-modal links, by presenting a series of experiments that study the connections in endogenous spatial orienting in hearing in vision. In these experiments, the spatial reorientation of the body was not required. Through seven experiments, subjects were required to determine the elevation of auditory or visual targets independent of their location or modality.

The results from the study showed that when subjects were conscious of the location of the stimuli, response times were reduced. This result was independent of the modality of the target. Also, when subjects were conscious of the modality of the target, a shift in attention occurred in the other modality.

This also resulted in shorter response times. In addition, when subjects were conscious of the cross-modal presentation of stimuli, the auditory and visual attention was commonly divided. These combined observations support the theory that endogenous covert spatial attention spatial attention does not only occur within a supramodal system. However, it also shows that the modalities do not occur independently either.

As a result, Spence and Driver suggest a new model for the division of attentional resources, which we refer to as the *separable but linked attentional system.*

Conclusions:

Understanding how attentional resources are divided across modalities is vital in the design of multimodal interfaces. The model suggested by Spence and Driver can aid designers in predicting the responses of operators to multisensory events, which in turn can assist in multimodal interface design. It should be noted that there are several models available, and thus more work is required in this area to determine which model is the best.

Reference:

Spence, C., & Ho, C. (2008). Multisensory warning signals for event perception and safe driving. *Theoretical Issues in Ergonomics Science*, 9(6), 523-554.

Overview:

This paper presents a review of design approaches for unimodal and multisensory warning signals used to alert drivers of potentially dangerous situations. Spence and Ho suggest new approaches to the design of multisensory warning signals, where the warning signals are presented in different regions of space surrounding the driver. This theory is critically examined using past research material.

The main finding of the research review is that stimuli which occur in the peripersonal space are processed differently than stimuli presented in the extrapersonal space. Those stimuli which occur in the peripersonal space are more demanding of attention. However, this concept presents a problem in that interface designers may want to relay information about event occurring in distant (extrapersonal) space. For example, interface designers may wish to alert the operator to a distant but urgent target, where the distance of the alert corresponds to the distance of the target – which makes use of a ecological valid representation of distance. However, the most effective warning signals are in the peripersonal space. Spence and Ho suggest that warning signals, or at least one component of them, should be presented in the peripersonal space. However, another component of the signal should be presented in the extrapersonal space, to communicate the spatial location of the event more accurately to the operator.

Conclusions:

The theory suggested by Spence and Ho indicates that warning signals should be placed in the peripersonal space. However, for portraying events which occur in the extrapersonal space, it may be useful to display a component of the warning signal in the extrapersonal space. This concept allows us to present warning signals more effectively in the design of multisensory interfaces.

Reference:

Van Rullen, R., & Koch, C. (2003). Competition and selection during visual processing of natural scenes and objects. Journal of vision, 3(1), 75-85.

Overview:

This paper presented a study to determine the number of objects that can be explicitly represented in one's short term visual memory. Van Rullen and Koch combined three paradigms called free recall, forced-choice recognition and visual priming to provide insights on the number of objects that access visual short term memory and whether objects in a visual scene were perceived even if subjects did not explicitly recall viewing them. Subjects were presented with a visual scene for 250ms that consisted of 10 objects and results showed that subjects could explicitly recall up to 4 objects with confidence and between 2-3 additional objects when asked to guess. The authors stated that there was a negative priming effect in regards to the objects that participants consistently failed to report. Negative priming occurs when a stimulus is presented in a visual scene eliciting "a trace of neural activity that can modify the processing of a subsequent repetition of the same/similar stimulus" which reflects "the suppression of ignored objects during attentional selection. This suggests that the ignored objects were represented in their visual system but was suppressed. Note that visual priming has shown that it is unaffected to low-level picture manipulations such as reflection but affects high-level properties of the stimulus such as semantic categorization

This paper describes the different capacities of the human visual system at different levels. It also tells us how negative priming effects can result in subjects not being able to recall certain objects at all. Thus, it is important to ensure that negative priming effects are eliminated/minimized. Further research is needed to examine how this recommendation can be transformed into an interface guideline. However, it is clear that interface designers must be careful using similar-looking objects in different displays since negative priming is possible.

Conclusions:

This study is also able to provide insight on how negative priming effects can result in mistakes in recalling objects. As stated in the paper, failure to recognize objects in a visual scene due to negative priming was caused by suppression of the objects because of other distracter objects. Thus, it is important to ensure that negative priming effects are minimized in interface designs. One possible solution to avoiding negative priming is to ensure that visual scenes or displays with a large amount of information and objects should be presented for longer durations so that users can better interpret and assimilate the information. An additional possible solution is to ensure objects/information within a visual scene is distinct.

Reference:

Vroomen, J., Bertelson, P., & de Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics*, 63(4), 651-659.

Overview:

This paper addresses the concept of a visual bias effect for simultaneous auditory and visual stimuli. Past research by the same authors indicated that the bias does not depend on the direction of endogenous attention. However, this paper instead addresses a similar concept instead concerning the direction of exogenous attention.

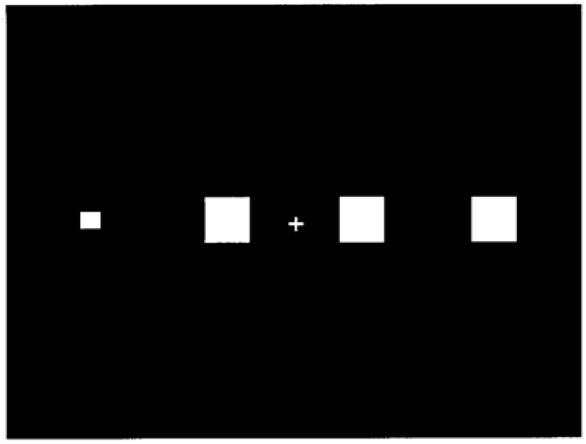


Figure A-40: An example of the stimuli used (p. 653)

One of the concepts required for the experimental study is that exogenous visual attention can be attracted towards a single object which is different in some dimension from all other items presented simultaneously (referred to by the researchers as a singleton). A visual display was utilized which presented four bright square with one square which was significantly smaller than the others. Three experiments were completed.

In the first experiment, subjects were required to make separate left-right responses to sound burst, which were presented simultaneously with singleton from visual modality. In the second experiment, subjects were asked to determine target letters presented either on the singleton or on the far-right large square. Lastly, the third experiment mixed the procedures of the second and first experiments to determine a control for potential differences in subjects' strategies in the other two experiments.

From the study, it was found that the obvious location of the sound was not attracted toward the singleton, but instead to the large squares on the opposite side of the display. Also, from the second and third experiments, it was found that performance decreased when the target was located on the large square as opposed to the singleton. This was compared with control trials where the singleton was not present, which showed support that the singleton attracted attention away from the target location.

From these results, it was concluded that visual bias of auditory sound location can be dissociated from exogenous visual attention. Also, the fact that a singleton can attract attention despite the fact that it is smaller than other items on the visual field is very important in interface design and signal presentation.

Conclusions:

The work presented in this paper supports the existence of cross-modal bias. Cross-modal bias is a serious issue that interface designers should take into account when designing multimodal interfaces. For example, the fact that a singleton can direct attention away from a target location can seriously affect the effectiveness of a display if the singleton is not expected.

<u>Reference:</u> Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: voluntary versus automatic allocation. *Journal of experimental psychology: Human Perception and Performance,* 16(1), 121-34.
<u>Overview:</u> In this study, an experiment was conducted to investigate whether an abrupt onset captures attention (called the abrupt onset effect). Subjects participated in a discrimination task in which they were presented with a display consisting of four letters arranged on the vertices of a hexagon and were told to discriminate which letters (either H or E) existed. On each trial, an arrowhead cue indicated the correct location of the target letter. The arrowhead cue's timing was manipulated by presenting it either 200 ms before, simultaneously, or 200 ms subsequent to the presentation of a test display (letter arrangement along the vertices of the hexagon). This paper cited other various research papers and stated that it has been established that subjects are capable of aligning their attention with a spatial location that contains task/goal related information within 200 ms of receiving that information (in this case the arrowhead indicating the target letter's location) (e.g. Eriksen & St. James, 1986; Murphy & Eriksen, 1987; Posner, 1980; Posner, Cohen, & Rafal, 1982; Remington & Pierce, 1984). Since the results in the onset and "no-onset" condition were identical, this study showed that precues resulted in highly focused attention and in return eliminated the abrupt onset effect in the discrimination task. Abrupt onsets seemed to only capture attention if the subject's attention was unfocused however as stated earlier, this was not the case when the subjects' attention was focused. If the distracter onsets captured the subjects' attention, then the performance should have varied in the no-onset condition compared to the onset condition. Thus, it was concluded that the abrupt onset effect is not valid if the individual is engaging in a highly focused attentional activity.
<u>Conclusions:</u> This study suggests that an abrupt onset will not capture one's attention if he/she is already engaged in an attentional activity therefore, this should interface designers should design the interface around this finding.

A.5 Intelligent Adaptive Interfaces

<u>Reference:</u> Hameed, S., & Sarter, N. (2009). Context-sensitive information presentation: Integrating adaptive and adaptable approaches to display design. In *Proceedings of the 53rd Annual Meeting of the Human Factors and Ergonomics Society* (pp. 1694-1698). Santa Monica, CA: Human Factors and Ergonomics Society.
<u>Overview:</u> Hameed and Sarter compare different adaptive forms of multisensory displays, describing the

downfalls and advantages of an adaptive, adaptable and hybrid displays. Adaptive displays are systems in which the interface is responsible for managing and organizing information presentation and task allocation. Although this form of adaptation will result in less continuous management of the interface by the user, it could also result in various implications. For example, since the interface controls automation, the operator's situation awareness will decrease and as a result performance may be affected negatively. This approach is employed by operators in a resistant manner due to the loss of situation awareness trend. Another form of adaption in displays is referred to as "adaptable displays". Adaptable displays systems in which the human operator manages and organizes information presentation in accordance to his/her preferences, interpretation of the system's status, context etc. Since the operator is in charge of adjustments, this approach solves the issue of reduced situation awareness in the previous approach however it also has very different downfalls. For example, the operator is required to manipulate the interface in accordance to various needs/preferences while simultaneously completing all his/her primary tasks, resulting in high workload and attention demands which could also result in performance decrements. A third approach is called "hybrid displays" which is a combination of both adaptive and adaptable displays in which the system and the user has the authority to share control of the interface. For example, the interface could perform actions and notify the operator of each automated action (solving the decreased situation awareness in adaptive displays) but at the same time the operator has the authority to intervene when he/she disagrees with the interface's choice of action (solves the work load and attention demand issue in adaptable displays). This approach combines the positive aspects of adaptive and adaptable displays while downplaying the disadvantages.

In addition to various adaptable displays, this article presents various adaptation drivers that are the underlying component of adaptive interfaces such as personal preferences, temporal demands, environmental conditions, user experience etc. The article also presents various methods to operationalize the operator's state and performance through electroencephalography (EEG) and event-related potentials (ERP) which is in concordance with other literature stating that EEG has demonstrated effective cognitive state classification.

Choice of modalities in adaptable interfaces could be determined by two factors: appropriateness and availability in relation to rank order values (0-1) indicating the modality's desirability level. This ranking system can also be applied to ambient/environmental conditions such as lighting, vibrations and sound. The ranking system is an example of a hybrid display because it incorporates the user's preferences and needs with interface automation but at the same time leaves leeway for the operator to intervene in modality choice if necessary.

<u>Conclusions:</u>

Hybrid displays appear to be the most effective since it appears to be in the middle of the interface having full authority and operator having full authority. It also serves as a middle ground between full automation and no automation.

<u>Reference:</u>

Hou, M., Gauthier, M. S., & Banbury, S. (2007a). Development of a generic design framework for intelligent adaptive systems. *Human-Computer Interaction*, 313-320.

Overview:

This paper explores guidelines for intelligent adaptive systems and developed a generic conceptual framework that consisted of the following four components that operate within a closed-loop system:

1. Situation assessment and support system – Consists of a real-time mission analysis, automation and decision support to provide information on the aircraft/vehicle/system's state and support the operator.
2. Operator state assessment – Real-time analysis of psychological, physiological and/or behaviour state of the operator within the context of a specific mission.
3. Adaptation engine – Utilizes the higher-order outputs from Operator State Assessment and Situation Assessment systems, as well as other relevant aircraft/vehicle/system data sources, to maximize the goodness of fit between aircraft/vehicle/system state, operator state, and the tactical assessments provided by the Situation Assessment system.
4. Operator Machine Interface (OMI) – The means by which the operator interacts with the aircraft/vehicle/system in order to satisfy mission tasks and goals and/or with the intelligent adaptive system, if applicable."

These are very good factors to take into consideration when developing an IAI since it allows the system to be able to absorb necessary information for all possible occurrences. Another framework referred to in this paper, consists of various models that is said to be incorporated together when designing IAIs. These components are as follows: "(1) *Organization Model*. This model incorporates knowledge relating to the organizational context that the knowledge-based system is intended to operate in (e.g., command and control (C2) structures, Intelligence Surveillance, Target Requisition and Reconnaissance - ISTAR etc.); (2) *Task Model*. This model incorporates knowledge relating to the tasks and functions undertaken by all agents, including the operator; (3) Agent Model. This model incorporates knowledge relating to the participants of the system (i.e., computer and human agents), as well as their roles and responsibilities; (4) *User Model*. This model incorporates knowledge of the human operator's abilities, needs and preferences; (5) *System Model*. This model incorporates knowledge of the system's abilities, needs, and the means by which it can assist the human operator (e.g., advice, automation, interface adaptation); (6) *World Model*. This model incorporates knowledge of the external world, such as physical (e.g., principles of flight controls), psychological (e.g., principles of human behavior under stress), or cultural (e.g., rules associated with tactics adopted by hostile forces); (7) *Dialogue/Communication Model*. This model incorporates knowledge of the manner in which communication takes place between the human operator and the system, and between the system agents themselves; (8) *Knowledge Model*. This model incorporates a detailed record of the knowledge required to perform the tasks that the system will be performing; and, (9) *Design Model*. This model comprises the hardware and software requirements related to the construction of the intelligent adaptive system. This model also specifies the means by which operator state is monitored."

Conclusions:

All the guidelines provided in this paper are very conceptual and as stated earlier, incorporate various variables of information that are necessary for an IAI to be able to adapt to the operator's needs and preferences effectively. Although this article states various concepts and areas to look at, it does not provide methods on how to operationalize such concepts. For example, how exactly is it possible to ensure that a system incorporates "knowledge" in relation to the tasks? What is meant by "knowledge"? What is considered relevant or irrelevant "knowledge"? The concepts stated above can result in vast amounts of information being included in those models. There seems to be an endless possibility of information the system may need. Set guidelines must be established to determine what exactly what information will be relevant and where the line will be drawn indicating that enough information has been provided to the system to result in an effective IAI.

<u>Reference:</u>

Hou, M., Kobierski, R. D., & Brown, M. (2007b). Intelligent adaptive interfaces for the control of multiple UAVs. *Journal of Cognitive Engineering and Decision Making*, 1(3), 327-362.

<u>Overview:</u>

This paper describes *Intelligent Adaptive Interfaces* (IAI) as "an operator interface that dynamically changes the display and/or control characteristics of human-machine systems to adaptively react to external events (mission and operator states) in real time. A typical IAI is driven by software agents (automation) that intelligently aid the decision-making and action requirements of operators under different levels of workload and task complexity by presenting the right information or action sequence proposal or performing actions in the correct format at the right time." The authors also state that IAIs will have the following capacities: the ability to model human decision making and control abilities, the capacity to monitor operator performance and workload and last but not least the ability to foresee the mission and/or operator's intentions. A key issue to address with IAIs are task allocation between the interface and operator. It is important to optimize "triggering conditions for task reallocation (e.g. by monitoring behaviour, cognitive states, physiological states, and situation events)." Task allocation amongst the operator and interface can significantly affect the operator's experience thus affect the overall mission and task performance due to potential issues in the automation domain such as task overload and decreased situation awareness. An additional criteria stated by this paper says that in order for the interface to be able to intelligently adapt to the operator and the mission goals, it is vital that information on the status of the system's and operator's goals are capable of freely flowing between both parties. The article outlines how the Defence Research & Development Canada (DRDC) conducted research projects in order to establish design guidelines for IAI systems. DRDC hypothesized that (1) IAIs will result in the operator's situation awareness increasing along with performance and a decrease in workload and (2) IAIs will be most effective in high workload situations. The experiment conducted by DRDC to test these hypothesises consisted of a scenario in which the Canadian Forces were assigned a task to provide security for a particular meeting. Authorities were then informed about a "lethal medium range UAV" and due to this UAV supposedly carrying plutonium or "dirty bomb" which could cause casualties and result in the region being inaccessible for many years if the bomb exploded, Canadian Forces must control the situation without firing at or attacking this suspected UAV. Various technologies were implemented in IAIs to test its effectiveness. Goals of this mission were established through a Hierarchical Goal Analysis (HGA) in which goals were presented and established in a hierarchical order from the highest level (e.g. highest goal = counterterrorism mission) to lower levels (e.g. goal = search sector level for other threats). This scenario was tested in two conditions, the first when the IAI condition was turned on and the second one when the IAI condition was turned off. Results depicted that operators had fewer task conflicts in the IAI ON condition (17%) compared to the IAI OFF condition (38%). In addition, there was a significant time reduction in the completion of high-level goals (more than 80%). Through these findings, the first hypothesis which was that IAIs would result in an increase of operator situation awareness and decrease their workload was shown to be valid.

In order to test the second hypothesis, DRDC presented the same synthesized scenario but to increase the workload and complexity of the scenario, multiagents were introduced had to communicate with each other in order to complete the assigned task effectively. Effective communication was essential for the mission to be successful. For example one agent would only be able to complete their portion of the task if they communicated the status of another agent's portion of the task, thus increasing workload. Performance was evaluated by 3 objective measures and 2 subjective measures. Objective measures included completion time and task shedding for critical task sequences (CTSs), an example of a CTS would be the time a UAV pilot took to control the UAV when the tactical navigator told him to do so. Additional objective measures were the number of airspace violations and trajectory along with a Situation Awareness Global Assessment Technique (SAGAT) score. The two subjective measures were perceived situation awareness and perceived workload determined by questionnaires. Results demonstrated that overall, the workload significantly decreased and situation awareness was increased in the IAI ON condition than in the IAI OFF. In addition, when IAI assistance was available, task shedding increased resulting in increased situation awareness regardless of the higher perceived workload. Therefore the second hypothesis regarding IAIs being the most effective in terms of situation awareness and performance in high workload situations is also valid.

Along with proving the effectiveness of IAIs, this paper briefly outlines design guidelines for IAIs. The main guidelines are as follows:

- The presence of HGA feedback as a display item. Operators should be allowed to return to a previous state of automation (an undo option). This appears to be an accurate guideline due to the concept's ability to serve as an effective scheme for goal organization.
- IAIs should inform the operator of any decisions or tasks it assumes or makes. Although this may seem like a good idea, the operator may not have enough attention resources or time to pay attention to all possible notifications therefore there needs to be an established guideline that determines to what extent the IAI informs the operator of its actions.
- The operator's state and intentions must be crystal clear and the IAIs perception of the operator's state and intentions should also be clear.
- An operator must have "suffice" trust in the IAI. An implication with this guideline is the lack of conciseness on how much trust is sufficient.

Conclusions:

Overall, this paper demonstrates the effectiveness of implementing IAI systems. However, during both experiments the IAI conditions were either completely off or on. There was no intermediate level of assistance provided to the operators where it may have been essential. This was an accepted constraint.

Reference: Maat, L., & Pantic, M. (2006). Gaze-X: Adaptive affective multimodal interface for single-user office scenarios. In *Proceedings of the 8th International Conference on Multimodal* Interfaces (pp. 171-178). New York, NY: Association for Computing Machinery.
Overview: This paper describes an existing multimodal IAI system called "Gaze-X" that was developed to support human-computer interaction. This interface models the user's actions and emotions and adapts the interface in accordance to the user's information and executes user-supported actions. Gaze-X can interpret various natural human communicative methods such as eye gaze direction, speech, facial expression, keystrokes and mouse movements. Gaze-X's reliance on human facial expression to interpret the user's mood can be problematic because this leaves leeway for conflicting interface interpretations. For example, if the user laughs when they are nervous due to fear, stress or tight deadline then the interface may interpret this as "happy" and not offer any assistance. This interface was designed for office tasks however specific types of tasks were not mentioned however the paper did state that Gaze-X's actions were case-based. This could be problematic in the UAV domain because assistance tends to be required when abnormal or less frequent occurrences such as emergencies occur but at the same time, implementing case-based adaptive assistance provided by the interface for this specific project in terms of the established used-case scenarios is a possible option. As mentioned in the IAI report component, Gaze-X coincides with design guidelines mentioned in other literature.
Conclusions: Gaze-X is an example of an adaptive multimodal interface. However, because Gaze-X focuses on very specific case-based scenarios, it may not be appropriate to support infrequent or abnormal situations that may occur during emergencies.

Reference: Meyer, B., Yakemovic, K., & Harris, M. (1993). Issues in practical application of an adaptive interface. In *Proceedings of the 1st International Conference on Intelligent User Interfaces* (pp. 251-254). New York, NY: Association for Computing Machinery.
Overview: This paper discusses various adaptation issues relevant to business environments that are not applicable to this project, but it also outlines various adaptation topics that are relevant. Meyer, Yakemovic and Harris states that an important aspect of designing an adaptable interface is determining what aspect of the system will adapt in response to an event. The authors stated the following as some ways a system can adapt: • "task allocation or partitioning – the system itself performs the complete task or part of it

- Interface transformation – the system adapts to make the task easier by changing the communication style and the content and form of displayed information
- Functionality – the system adapts the functions available to each user
- User – the system can help the user to adapt by determining apparent problem areas and providing intelligent tutoring for them."

The most effective form of adaptation depends on the type of task and even part of the task. This paper also outlines criteria for determining when adaptation should occur which include "user experience, aptitudes, demographics, task complexity and frequency, probable workload and physical conditions." These criterions coincide with other literature and take all the necessary factors into consideration. An interesting way to obtain data on workload mentioned in this paper that has not been mentioned in other literature is that the adaptive system can measure and record the time it takes the user to accomplish tasks and relevant subtasks and adjust the level of assistance in accordance to an "expertise level" proportionate to the recorded speed. This appears to be a possible less invasive approach to obtain workload measures compared to EEG and other intrusive methods.

Conclusions:

One of the most important aspects of designing an adaptive interface is in determining what portions of the system will "adapt" and respond to changes in the environment or user. Task allocation, interface transformation, functionality, and changing the user are all possible candidates.

Reference:

Pentland, A., & Roy, D. (1998). Multimodal adaptive interfaces. In *Papers from the 1998 AAAI Spring Symposium on Intelligent Environments* (pp. 115-122). Menlo Park, CA: Association for the Advancement of Artificial Intelligence.

Overview:

Pentland and Roy created a human machine interface that is "centered around on-line learning to actively acquire communication primitives from interactions with the user" by utilizing natural modalities such as speech, hand gestures and vision. The user can interact with the system through a synthesized animated character referred as "Toco the Toucan." The reason for this interface being centered around on-line learning is because the authors wanted to solve the reference resolution problem which is the process of "inferring the user's intent based on observing his/her actions." Implications can arise when a user refers to something like "the button" in which the interface may not know which "button" the user is referring to. Thus, the authors concluded that in order to make an effective adaptable interface, that is aware of the specific words and gestures the user uses along with their intent, a speech recognizer should be capable of learning a wide variety of vocabulary. The design behind this interface solves the reference resolution problem because the user basically teaches the interface associations between words and meanings. For example if the user points to an object and says "Toco, button," the interface will then associate that object with the word "button" and eventually build complex schemas. This design has possible downfalls such as time constraints along with conflicting word schemas. This would take a long period of time for "Toco" to learn all necessary objects and commands to perform a precise action. In addition, if the user presents conflicting words to "Toco," this could result in implications. This system detects the user's hand gestures through a vision system consisting of colour video cameras capable of sensing human hand gestures. Toco the Toucan meets various adaptive multimodal design guidelines mentioned in previous literature such as the interface being capable of informing the user of its interpretation/perception of his/her requests/commands. Toco is capable of modeling his behaviour state by depicting various behaviours that display his attention state and interpretation. For example, if the user has caught Toco's attention, Toco's eyes will widen and he will look in the direction of the object he interpreted the user referring to. It is important to note that this interface is in its preliminary design stages and future work is planned to improve its dynamic task ability. For example, assigning Toco to perform actions on a specific object.

Conclusions:

Toco the Toucan can serve as an example of how it is possible for humans and interfaces to communicate in a very natural way and perhaps this is the key to increase trust and reliability amongst users and the system. It is important to note that a lot of the literature on adaptable interfaces are attempting or executing adaptable multimodal interfaces via natural human methods of communication such as speech, vision and hand gestures. This is due to the large amount of research that states that communication should be in the most natural and comfortable form for an effective adaptive interface.

Reference:

Reeves, L. M., Lai, J., Larson, J. A., Oviatt, S., Balaji, T. S., Buisine, S . . . Wang, Q. Y. (2004, January). Guidelines for multimodal user interface design. *Communications of the ACM*, 47(1), 57-59.

Overview:

This paper provides a brief outline on general design guidelines for multimodal interfaces. For example, Larson, Reeves and Oviatt state that human cognitive and physical abilities should be maximized by avoiding unnecessary presentation of information in different modalities when the information must be attended to simultaneously. Factors like the user's memory should also be maximized by using complimentary modality combinations such as "system visual presentation being coupled with user manual input for spatial information and parallel processing along with coupling auditory presentation with user speech input for state information, serial processing, attention alerting or issuing commands." Other important aspects of modality combination are that users should be able to choose between modalities and that the system is able to capture the user's interaction history so that it can record user preferences (adaptivity).

Conclusions:

This paper presented very broad guidelines, lacking specifics. For example, it stated that compatible modality combinations should be programmed, but did not mention in detail compatible modality guidelines. Although this article provided basic guidelines, it does not appear to be very useful in terms of specific design guidelines.

Reference:

Schneider-Hufschmidt, M., Groh, L., Perrin, P., Hine, N., & Furner, S. (2003). Human Factors guidelines for multimodal interaction, communication and navigation. In *Proceedings of the 19th International Symposium on Human Factors in Telecommunication*.

Overview:

This paper discusses issues and solutions for multimodal interaction and presents design and implementation principles. Although this paper is more geared for multimodal interaction for the disabled, the authors' design principles can be applied to a broad use of multimodal interfaces. The following guidelines provided within this paper are as follows:

- "Use multimodal presentation of information to allow users with different preferences and abilities to interpret data in their preferred way." This is where the adaptive component of multimodal interfaces comes into play. It is important that the interface is able to take into account personal preferences along with environmental conditions and is able to do this automatically. The authors also pointed out that the modalities should be "scalable" in which users have the option to adjust individual modalities (i.e. display contrast, audio level) to suit their needs.
- "It should be possible to choose different presentation modalities using any of the available interaction modalities." This could help the user modify/adjust any automation of information presentation that the interface selected.
- "The user-specific modality setting should persist."
- "The same information should be expressed in different modalities." This is a good point because it is important that users can access the same exact information across modalities and that this information is stored in "delivery-independent form"

Conclusions:

This paper provides basic rules of thumb that designers should consider when developing adaptive multimodal interfaces. It advocates providing the user with choices through which modality information is presented in, while still stating that redundancy should be provided by having the same information provided through different modalities.

Reference:

Tripathi, P. (2008). Human-Centric Framework for Perceptually Adaptive Interfaces. *Framework*, 255-256.

Overview:

This paper attempts to provide a conceptual framework for the design and development of multimodal IAI in terms of interaction between human and computer. The authors approached this issue through two main research questions (1) "how does the perceiver integrate information about features from sensory modalities and (2) how does multisensory integration affect performance?" and concluded that the following questions must be addressed in the early stages of multimodal adaptive interfaces:

1. "Choice of the information that is to be conveyed ("content selection").
2. Selection of modalities through which the information will be conveyed ("modality allocation").
3. Selection of the format in which the modalities will be able to perceive that information ("modality realization").
4. Determinations of mechanism(s) that are used combine the modalities ("modality combination").
5. Evaluating the affect of environmental and cognitive factors on user's perceptual integration ("Situated multimodality").
6. Analysis of performance of the human user in the interface ("Task Analysis")."
7. These are all very important aspects that must be considered prior to developing the interface.

Conclusions:

This paper provides some general guidelines for the design of adaptive interfaces, and explicitly discusses how modality can be used as one method of adapting the interface. This paper states that modality allocation and modality realization are both choices that have to be made when designing an adaptive multimodal interfaces.

Reference:

Sherry, R. R., & Ritter, F. E. (2002). *Dynamic Task Allocation: Issues for Implementing Adaptive Intelligent Automation* (Report No. ACS 2002-2). University Park, PA: The Pennsylvania State University School of Information Sciences and Technology.

Overview:

Sherry and Ritter conducted extensive research about various types of automation and its usability and provided recommendations in regards to improving pilot/automation task allocation which are as follows:

- Avoidance of multiple options – research has suggested that when humans must make decisions in "real-world time constrained" situations, we tend to make decisions based on existing schemas of personal experience and training therefore, if the interface provides multiple options, the operator may get overwhelmed and as a result cause. This coincides with other guidelines provided. The interface should be able to only provide the most relevant options and not overwhelm the user with numerous unnecessary options.
- Minimize interruptions – under time constrained conditions, a task interruption can result the user making errors thus IAIs should be context-sensitive and be able to effectively prioritize tasks and interruptions.
- Operators should be an active participant – research has shown that humans perform poorly on tasks that require continuous monitoring therefore they should be actively

engaged in participation so factors like loss of situational awareness and vigilance are not an issue

- Humans must be given control authority – an issue with IAIs are whether humans or the interface should be given the authority to make decisions. This paper states that since humans are held responsible for the overall task, they should be given control authority. This may lead to an increased workload on the human's part however if a fine balance of task allocation is achieved, issues like overload can be avoided. Previous literature stated that a hybrid approach to adaptation and automation holds a middle ground for control authority and automation between the operator and interface which seems to be effective.
- It is important that the interface is capable of clearly indicating its behaviour and state. This is an important design feature that can assist with tackling the issue many researchers have brought up in which automation causes decrements in situation awareness. If the interface is able to notify the operator of its behaviour and state but do so in a way where this is not distracting or redundant, automation will not result in a decrease of situation awareness.
- Research has suggested that intermediate levels of automation may be optimal. According to Sherry and Ritter, research has shown that the task implementation assistance (automation) results in negative overall task performance in higher level cognitive functions such as decision making. Since an interface is not capable of exactly mimicking a human's complex mind, it appears that humans should perform high level of cognitive functions.

Conclusions:

This research recommends what steps are needed to improve pilot/automation task allocations. Avoiding situations where multiple options are provided allows the pilot to make decisions without being overwhelmed. Care must also be taken to avoid interruption of the pilot so that tasks are carried out without errors. Control authority should also be given to the pilot as this allows the pilot to be an active participant in the automated task. It is also important that the interface is clear for the pilot to understand so that it is possible to optimize their task allocations.

A.6 Developing a Program of Research

Reference:

Aretz, D., Andre, T., Self, B., & Brenaman, C. (2006). Effect of tactile feedback on unmanned aerial vehicle landings. In *Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC)* (Vol. 2006).

Overview:

In this paper, Aretz et al. explore if providing tactile feedback to UAV operators during training for a landing task improves performance. Tactile feedback was provided via a tactile vest with four rows of tactors. Each of the rows represented different levels of deviation from the optimal altitude during the approach. The top most row would vibrate intensely (200ms on, 100ms off) if the UAV was 20 feet above the optimal glideslope. The second highest row would vibrate softly (100ms on, 600ms off) when the UAV was 10 feet above the optimal glideslope. A similar coding strategy was used for the bottom two rows for when the UAV was below the optimal glideslope. The authors hypothesized that the duration of training required until successful performance was achieved would be decreased with the use of vest feedback. They also predicted that the differences in glideslope RMS between the training phase and a post-training phase (where participant's switched from using the vest to not using the vest or vice versa) would be smaller for participants who were trained without the vest and subsequently had to fly the post-training trials with the vest than for those who received initially received training with a vest.

Methodology:
Participants interacted with the UAV simulation using a throttle control and stick. The participants were also provided with two visual displays: one of a camera view from the nose-mounted camera, and one with a map and other mission relevant data. The experiment had one independent factor, whether they had training with or without the tactile vest feedback. Prior to the main experimental task, participants were trained on how to use the simulator and read the displays using a flight manoeuvring test.
The primary experimental task was the UAV landing task. Participants were required to manually fly the UAV from downwind of the airfield to its final landed position. To simulate training, participants were required to repeat the landing task until they received a passing score (a RMS error of 20 feet or less). The number of attempts required until the participant was able to obtain a passing score was recorded as a dependent variable. The glideslope RMS error during the training was also used as a dependent variable; however the paper did not specify how this RMS was calculated. Most likely, this was an average RMS value over all the pre-passing trials.
After achieving a pass-score, the participants were required to complete three more landings. However, participants who had used the vest during training were required to fly without the use of the vest, while participants who were trained without the vest were provided the vest for these landings. The glideslope RMS for the post-trial landings were also used as a dependent variable.

Results:

Table A-6: Performance results of participants with vest and without vest

		Vest	No Vest
# of Trials	*Mean*	3.53	7.13
	SD	1.41	2.72
Glideslope RMS Error	*Mean*	38.57	54.53
	SD	12.18	23.82
Post-trial Glideslope RMS Error	*Mean*	31.04	25.82
	SD	10.40	10.49

The previous table presents the performance of participants by condition. Participants who had tactile feedback were able to achieve passing scores much faster than those who did not have the vest, and the average glideslope RMS during the training was much larger for the no vest condition. This supported the author's primary hypothesis. However, there were no differences between the post-training RMS values. A two-level interaction between pre/post training and vest condition (vest or not vest) was found. While both groups improved performance due to the training received, those who received training without tactile feedback experienced a larger increase in performance (decreased RMS) when compared to those who received training with the tactile vest. The authors concluded that "the vest group had probably experienced most of their learning while initially wearing the vest whereas the no-vest group was still learning during the initial trials without the vest. Switching to the vest significantly added to their learning performance and dramatically decreased their RMS error." It is interesting to note that the post-trail RMS errors were all below the passing grade required to progress past the training.

Conclusions:

Tactile feedback can affect how quickly an operator is able to achieve good performance in a task. Even when operators are trained without tactile feedback, a correctly designed tactile feedback system can increase performance. The tactile vest used in this study had an interesting coding method that used both spatial location and intensity as a way of showing error. The changes in intensity were a good way of modulating saliency and urgency of the tactile cue while the spatial location worked as a redundant cue of saliency but provided a intuitive mapping of the error while also providing context information (is the error due to being above or below the optimal glideslope).

Reference:

Brill, J. C., Mouloua, M., Gilson, R. D., Rinalducci, E. J., & Kennedy, R. S. (2008). Effects of secondary loading task modality on attentional reserve capacity. In *Proceedings of the 52nd Annual Meeting of the Human Factors and Ergonomics Society* (pp. 1219-1223). Santa Monica, CA: Human Factors and Ergonomics Society.

Overview:

This paper attempts to measure the reserve attentional capacities of vision, audition, and the tactile modality using a secondary loading task. The Multi-Sensory Assessment Protocol (M-SWAP) is a secondary task measure which makes use of perceptual signals in different modalities to gauge reserve cognitive capacity. The authors were interested in testing if there was evidence for different resource pools for each modality, as suggested by Wicken's *Multiple Resource Theory* (MRT). To test this, a primary visual monitoring task was used to load the visual modality, while the secondary task measured reserve cognitive capacities in the visual, auditory and tactile modalities.

Methodology

The primary experimental task used the Multi-Attribute Task Battery (MATB) created by Comstock & Arnegard (1992). The MATB is primarily a visual task, consisting of four horizontally arranged bars with a moving pointer. The bars represent pressure and temperature readings from aircraft engines. During the course of the experiment the participants were required to monitor for "malfunctions" in the gauge readings. Once the participant detected a malfunction, they had to respond by "resetting" the gauge using the keyboard.
The M-SWAP was used as a secondary loading task to assess the reserve cognitive capacity of the participants while engaged in the mainly visual primary task. The M-SWAP secondary task

requires that participants count the number of signal presentations in a particular information channel during the course of the experiment. Each modality consisted of 3 possible channels of information. In the visual modality this was represented by three white boxes which could be turned on or off, in the auditory modality this was three tones at different frequencies, and no information was listed for how the three channels in the tactile dimension were separated. Participants were asked to perform the MATB task. Over the course of the experiment, the M-SWAP task was presented at different times.

<u>Results</u>

No significant differences were found for the primary MATB task, which indicates that participants were indeed treating it as a priority across all secondary task conditions. However, differences were found for the M-SWAP secondary task. The hypothesis that visual counting performance in the secondary task was lower compared to the auditory and tactile conditions was supported. Brill et al. also found that workload (as measured by NASA-TLX scores) was significantly higher in the visual counting condition, than in the non-visual conditions.

Taken together, the authors reasoned that two possible explanations could account for the results. The first is one that is consistent with MRT, and that each modality had its own pool of resources. Thus, when the visual modality was loaded using the MATB, performance in the secondary visual task would decrease because there would be fewer resources available. The second possible explanation was that there was a single resource pool for all modalities, and that each of the tasks "consumed approximately the same quantity of resources, as they imposed comparable levels of demand." The authors favour the MRT explanation because of trends in their data.

One possible explanation not taken into account by the authors is that the bottleneck in the visual secondary task was due to properties of the sensory organ and not attentional resources. The paper did not describe whether the secondary visual task required overt attention orientation. Also, the paper did not consider other possible explanations from different models of Multimodal attention (such as the independent but connected model advocated by Spence and Driver (1996)).

<u>Conclusions:</u>

M-SWAP is a potential secondary task that can measure loading in different modalities. Using a visual primary task reduced visual counting performance in the M-SWAP secondary task. Attention capacities for audition and touch are similar. Findings suggest that there are relatively independent resource pools for each modality.

<u>Reference:</u>

Calhoun, G., Draper, M., Ruff, H., Fontejon, J., & Guilfoos, B. (2003). Evaluation of tactile alerts for control station operation. In *Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 2118-2122). Santa Monica, CA: Human Factors and Ergonomics Society.

<u>Overview:</u>

This paper evaluates the use of tactile alerts in a simulated UAV GCS task. Two tactors, one located on each wrist, were used as non-spatial alert cues. Three different UAV system faults were mapped onto the tactors. A 2x2 design of alert condition (Baseline vs. Tactile) and mission difficulty (Easy vs. Difficult) was used.

Methodology

The experimental task involved a tracking/flight navigation task, where the participants were asked to maintain an altitude and airspeed while flying along a path in the UAV simulator, and “check list” tasks, where the participants had to respond to an alert and follow a series of data input steps. There were 5 different types of check list tasks, though only three of them were accompanied by an alert: non-critical warnings, critical warnings, and information queries. The other two check list tasks were routine navigation or waypoint update tasks. For the three alerts, participants were required to make a response which confirmed the detection of the alert, and then they were required to complete the data input steps required. Information query alerts had an additional step where the operator had to respond to a visual stimulus within 10 seconds, otherwise the alert was counted as a miss. Different alerts were used for the baseline and tactile conditions which are shown in the following table. Each participant was given four hours of training, and care was taken to ensure that participants could reliability perform the tasks required (both independently and concurrently).

Table A-7: Experiment Conditions for Calhoun et al. (2003)

TASK	DIFFICULTY		ALERT CONDITION	
	LOW	HIGH	BASELINE	TACTILE
UAV Turns	1	3		
Normal Operations	2	4	No alerts for routine in-flight and waypoint update tasks	
Non-Critical Warnings	1	3	*Visual*: Colored “A” or “D” on HUD *Auditory*: Yes*	*Visual*: Colored “A” or “D” on HUD *Auditory*: Yes* *Tactile*: None; tactile reserved for Critical Warnings & Information Queries
Critical Warnings	3	3	*Visual*: Colored “A” or “D” on HUD *Auditory*: Yes*	*Visual*: Colored “A” or “D” on HUD *Auditory*: Yes* *Tactile*: One tactor vibrated: left-arm tactor: icing right-arm tactor: servo overheat
Information Queries	2	2	*Visual*: Red “QUERY” on HUD *Auditory*: None	*Visual*: None *Auditory*: None *Tactile*: Both tactors vibrated
*Auditory Alert: 0.4 second klaxon sound, primarily at 487 Hz; 10.3 dB.				

Results

Reaction time between onset of alert and participant response was used as a measure of the effectiveness of the different alerts. A significant effect of alert type (Baseline vs. Tactile) was found for information queries, where the tactile condition produced faster responses than the baseline condition. None of the other effects were found to be significant. Calhoun et al. suggest that this may have been because two tactors were used for the information query alert, while only one tactor was used for the warnings. In addition, the information query tactile alert was entirely tactile while the baseline condition was entirely auditory. The warning alerts featured multisensory alerts that used vision, audition, and touch in the tactile condition. The authors conclude that tactile cues, in the presence of auditory and visual cues did not hinder or improve performance. However, uni-modal tactile cues (the tactile information queries condition) did produce faster response times than uni-modal visual cues (the baseline information queries condition).

The participants' performance during the tracking/flight task was also analyzed, however no significant effects of alert types were found. This led the authors to conclude that the introduction of the tactile cue did not allow the participants to direct more resources to the flight task.

Conclusions:

This experiment uses checklists as a secondary task to increase the workload of the participants. Tactile cues, when present with auditory and visual cues did not improve or degrade performance. There is some evidence that uni-modal omni-directional tactile cues work well as non-redundant cues for alerts.

Reference:

Calhoun, G., Fontejon, J., Draper, M., Ruff, H., & Guilfoos, B. (2004). Tactile versus aural redundant alert cues for UAV control applications. In *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 137-141). Santa Monica, CA: Human Factors and Ergonomics Society.

Overview:

This paper evaluates the ability of redundant aural alerts and redundant tactile alerts to improve performance in a simulated UAV GCS task. Two tactors, one located on each wrist, were used as non-spatial alert cues. Three different UAV system faults were mapped onto the tactors. A 3 x 2 design was used with alert condition (Baseline vs. Redundant Aural vs. Redundant Tactile) and auditory load (Heavy vs. Low). Only experiment 2 is reported, but results from experiment 1 are similar.

Experiment 2
Methodology

Table A-8: Experiment Conditions for Calhoun et al. (2004)

ALERT CONDITIONS		Visual	Redundant	
			Aural Cue	Tactile
Baseline	**Critical**	Red “C” on HUD & red HDD text	none	none
+Aural	**Critical**	Same as Baseline	Type 2	none
+Tactile	**Critical**	Same as Baseline	none	wrists

The experimental task involved four concurrent tasks: a tracking/flight navigation task, a warning response data entry task, a radio frequency data entry task, and an IFF task. The navigation task was similar to the one used in Calhoun et al. (2003). Participants were asked to maintain an altitude and airspeed while flying along a path in the UAV simulator. In the warning response data entry task, participants responded to critical alerts (which differed in presentation based on alert condition). After responding to the alert, participants were also asked to follow through a series of steps related to the alert issued. The alert conditions used are shown below. The saliency of the aural and tactile cues was found to be equivalent in a pre-test.

The radio frequency data entry task was based on the Coordinate Response Measure used by Bolia, Nelson, Ericson, and Simpson (2000). Radio calls, composed of a call sign, a colour and a number (e.g. ready *Eagle,* go to *blue 8)*, were played, and participants were required to respond to a specific call sign and conduct a data entry task based on the colour and number in the radio call. The auditory load was manipulated by having only relevant calls sign for the Low auditory load condition, and by having 8 different call signs for the High auditory load condition. Finally, the IFF task required that participants respond to visual stimuli using radio messages.

<u>Results</u>
Reaction time between onset of alert and participant response was used as a measure of the effectiveness of the different alerts. A significant effect of alert type (Baseline vs. 2nd Aural vs. Tactile) was found: the baseline alert type was significantly slower than the 2nd aural alert and the tactile alert. The participants’ performance during the flight navigation task was found to be worse in the baseline condition than the tactile condition. Auditory load had the expected effects, with high auditory load resulting in a higher amount of perceived workload and task difficulty. Participants were also able to complete more radio tasks in the Low load condition compared to the High load condition.

<u>Conclusions</u>
The authors concluded that redundant non-visual alerts improved performance over using just the visual alerts. They also found that aural alerts were just as effective as tactile alerts even in varying conditions of auditory load. It is also worth noting that the baseline condition was only significantly different from the other two alert conditions in the high auditory load condition in experiment 1.

Conclusions:

The concurrent radio task could be used to increase auditory load (however, a manipulation check is required). Tactile and aural redundant cues both improve performance.

Reference:

Donmez, B., Graham, H., & Cummings, M. (2008). *Assessing the Impact of Haptic Peripheral Displays for UAV Operators* (Report No. HAL2008-02). Cambridge, MA: MIT Humans and Automation Laboratory. Retrieved from http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA479798&Location=U2&doc=GetTRDoc.pdf

Overview:

This paper examines the effectiveness of continuous and discrete haptic peripheral displays in a UAV supervisor control scenario. The haptic display was used as a redundant cue to a visual display which showed the location of multiple UAVs as well as a scheduling/timeline tool that helped the operator decide when UAVs were able to deploy their payload to targets of different priorities. The experiment made use of the Multiple Autonomous Unmanned Vehicle Experimental (MAUVE) test bed and stimuli were displayed using a Multimodal workstation (MMWS) featuring a multi-monitor visual display, over-the-head headset for auditory information, and an inflatable pressure vest and vibrating wristbands for haptic information. The experiment had a single factor, haptic feedback type (continuous vs. threshold). Participants also received information about two different variables using the haptic feedback: late target arrivals and course deviations.

Continuous feedback for late target arrivals was displayed by inflating the vest to varying degrees based on the priority of the target (this was manipulated by inflating a greater number of air bladders in the vest for high priority targets and less for medium and low priority targets). The vest stayed inflated until either the operator responded to the late arrival or the UAV continued onto the next target. Continuous feedback for course deviations was provided by buzzing of the wristband. As the course deviations became larger, the buzzing intensified by increasing the number of activated motors and decrease the time between activations of the motor.
Threshold feedback for late target arrivals was displayed by inflating the vest for a 2000ms interval when the late arrival was detected by the scheduling tool. Threshold feedback for course deviation was displayed by buzzing the wristband at full intensity for 600 ms when the UAV deviated from its course by 10 degrees.

Experimental Task
The primary experimental task was a UAV supervisory control task. The participants were required to monitor four UAVs while correcting for course deviations and making decisions about whether to pursue targets based on the scheduling information provided by the decision support aid. Participants could correct for course deviations by clicking on a reset navigation button. To correct for late arrivals, participants were asked to skip low priority targets while making a decision of whether to skip or request delays for medium and high priority targets based on information provided by a decision support visualization tool. The participants were also required to respond to an auditory secondary workload task. This task involved responding to air traffic

control radio chatter when the word “Push” was heard.

Results

Continuous haptic feedback was found to produce significantly faster response times for course deviations, while threshold haptic feedback produced faster response times for late arrivals. No significant effects of feedback type were found for the secondary auditory loading task and for subjective measures of workload (as measured by NASA-TLX). Taken together, the authors suggest that continuous information, such as the deviation from the course, is best supported using continuous feedback, while discrete events, such as a late arrival, are best supported using threshold feedback. The post-test feedback also revealed that participants liked the threshold feedback for course deviations more than the continuous feedback even though the threshold feedback produced slower reaction times. The authors state that this mismatch between the subjective and actual performance could be due to participants being annoyed by the continual buzzing of the wristband.

Conclusions:

Continuous feedback should be used for continuous data. Threshold feedback should be used for discrete events. Participants may have mismatches between cues that they like and their actual performance using the cue. There did not seem to be any differences in the ability to handle a secondary task between continuous and threshold feedback.

Reference:

Kramer, L. J., & Busquets, A. M. (2000). *Comparison of Pilots' Situational Awareness While Monitoring Autoland Approaches Using Conventional and Advanced Flight Display Formats* (Report No. NASA-2000-tp210284). Hampton, VA: Langley Research Center. Retrieved from http://portal.acm.org/citation.cfm?id=887327

Overview:

This paper describes the evaluation of three advanced autoland displays for commercial aircraft. The focus of this summary will be on the situation awareness methodologies used to evaluate the different designs. Situation awareness (SA) measures and workload measures were used to gauge the effectiveness of each display. Participants were asked to monitor autoland operations while using one of four display concepts (one baseline display, and three advanced interfaces). Scenarios, such as conflicting traffic situation assessments, main display failures, and navigation/autopilot system errors were used to measure the pilots’ situation awareness and workload.

Three SA measures were used:

- *Anomalous Cue/Detection Time Technique*: This method measures the time between introduction of a problem or fault and its detection, diagnosis, and response. Specific scenarios must be designed that allow for problems to be introduced.
- *Freezing/Probes*: In this method the experimenter either “interrupts a task or ‘freezes’ the

task and then proceeds to take some form of measurement". Many different probes can be used during a freeze including asking about the current state of objects or variables, or about future events.
- *Subjective Methods*: Questionnaires that ask for feedback on the subject of situation awareness.

Two workload measures were used:
- *Modified Cooper-Harper Ratings* (see Donmez, Brzezinski, Graham and Cummings, 2008 for an implementation of modified Cooper-Harper Ratings for Unmanned Vehicle Displays).
- *Subjective Methods*

Experimental Task

The primary experimental task was to monitor the aircraft interface as it performed an autolanding during a standard approach. The experimenter used 11 types of experimental scenarios: normal run, flight director conflict with autopilot, flight director conflict with raw data, aircraft incursion on final, flag take-off/go-around, two navigation system error scenarios, three blanking scenarios, and a probe approach scenario. These scenarios were designed to help probe for differences in SA. Normal scenarios were randomly distributed in the experiment to reduce the participant's expectations of abnormal events.
Six of the scenarios were evaluated using the anomalous cue/detect time technique (flight director conflict with autopilot, flight director conflict with raw data, aircraft incursion on final, flag take-off/go-around, two navigation system error scenarios). These scenarios introduced some sort of problem to the landing, which the participants had to detect and correct for. The remaining four scenarios were evaluated using the freezing/probe technique (three blanking, and one probe approach). In the blanking scenarios, display system failures were simulated by blanking the screen, and participants had to manually fly the rest of the approach using a single backup instrument. The probe approach froze the simulation and probed the participant with a series of questions related to SA. Prior to the blanks and freezes, deviations from nominal autoland behaviour were introduced.

The participant was able to press two buttons during the scenarios. The first button, labelled 'CONCERN', was pressed to indicate that the participant had detected a fault. The second button, labelled 'TOGA' (Turn-Off/Go-Around), was pressed when the participant felt that the autopilot should be disconnected and that manual flight was required.

Dependent Measures

Anomalous Cue/Detection Time Scenarios
- Detection time: time from introduction of problem until the 'CONCERNED' button was pressed.
- Reaction time: time from introduction of problem until the 'TOGA' button was pressed.
- Difference between detection time and reaction time.

Blanking Scenarios
- Vertical path error RMS, mean, and standard deviation.
- Lateral path error RMS, mean, and standard deviation.

- Distance from path RMS, mean, and standard deviation.

Probe Scenario

- Subjective questions about their situation awareness.
- Whether they detected the abnormal flight conditions.
- The authors noted that the participants reported that the 'surprise' of having a probe scenario (which was different from the other scenarios) caused them to remain more vigilant in the following trials.

Subjective Questionnaires

- Six questionnaires were used (one for each display concept, one comparing the three display concepts, and one for the probe scenario).

Analysis Techniques

Anomalous Cue/Detection Time Scenarios: Repeated measures ANOVAs were used to test the differences between the different displays for each of the metrics (detection time, reaction time, difference between detection and reaction time).
Blanking Scenarios: Repeated measures ANOVAs were used to test the differences between the different displays for each of the metrics (path error RMS, means, and standard deviations).
Probe Scenario: A count of the number of abnormal flight conditions detected.
Questionnaires: ANOVAs on the ratings.

Conclusions:

An excellent resource for our project. This paper has a very similar design problem as our current project (interfaces for autoland scenarios). The SA measures used and the procedures can be replicated for our experiment. SA measurements can provide insights into how an operator uses an interface that may not appear in performance based metrics.

Reference:

Maza, I., Caballero, F., Molina, R., Peña, N., & Ollero, a. (2009). Multimodal interface technologies for UAV ground control stations. *Journal of Intelligent and Robotic Systems*, 57(1-4), 371-391.

Overview:

Maza et al. examines different types of multimodal input and output technologies the context of UAV ground control stations (GCS). The authors describe two flows of information between the GCS and the operator. Multimodal presentation can be used for information flowing from the GCS to the operator. These include:

- 3D audio
- Speech synthesis
- Haptic devices

Information flow from the operator of the GCS can also be mediated by multimodal technologies such as:

- Touch screens

- Automatic speech recognition
- Operator's state
- Head tracking (2DoF or 6DoF)
- Eye tracking (2DoF)
- Body motion sensors

A simple multimodal input/output testing task was created where each trial contained either a yes button or a no button located at a random location on a display. Participants were required to press the yes button while allowing for no buttons to time-out, and they were required to do this as quickly as possible. A number of different conditions were tested using different multimodal input and output combinations as shown in the figure below.

Table A-9: Tests using different multimodal input and output combinations

Experiment nr.	Description
#1	Mouse interface only
#2	Touch screen interface only
#3	Touch screen and speech synthesis
#4	Touch screen and 3D audio
#5	Touch screen and tactile interfaces
#6	Touch screen, 3D audio and tactile interfaces
#7	Touch screen interface test repetition

Accuracy was very high for each of the conditions, so reaction time information was analyzed. Probability density functions were calculated for each condition, based on the data gathered. The authors found that with each successive application of a new multimodal technology, reaction times decreased. The analysis method used in this paper (probability density functions to represent reaction times) is uncommon within the literature. One possible motivation for using this kind of analysis was due to a low sample size (9), while having a large number of trials (363-381). However, this did not allow for an easy statistical test of the differences between the conditions. Therefore, the results between the different technologies seem to be based on qualitative judgements of the probability density functions.

Conclusions:

Multiple redundant multimodal presentations can increase reaction time. Touch screens increase the response time when compared to mouse only interfaces. This study used an unusual analysis strategy.

Reference:

Oskarsson, P., Eriksson, L., Lif, P., Lindahl, B., & Hedström, J. (2008). Multimodal threat cueing in simulated combat vehicle. In *Proceedings of the 52nd Annual Meeting of the Human Factors and Ergonomics Society* (pp. 1287-1291). Santa Monica, CA: Human Factors and Ergonomics Society.

Overview:

This paper describes the evaluation of three multimodal threat cueing displays for a simulated combat vehicle. The focus of this summary will be on the methodology used. Previous research had shown that auditory and tactile cues, used in conjunction with a visual cue could help orient an operator's attention. However, much of the previous research had been done with heads-down displays. Oskarsson et al. hypothesized that providing the visual cue on a head-up display (HUD) would also provide similar advantages to threat orientation, by providing better localization.

Methodology

Three multimodal interfaces were tested: a HUD + 3D audio display, 3D audio + tactile belt, and a HUD + 3D audio + tactile belt display. The visual cue was composed of a rectangle with overlaid arrows pointing in the direction of the threat. The 3D audio was presented through headphones, and a head- tracker compensated for head movements. The tactile belt consisted of twelve tactors distributed equally around the belt. Each tactor covered a 30 degree sector of the horizontal dimension.

The primary experiment task for the participants was to drive the simulated combat vehicle along a road until a threat occurred. When the threat appears, the participant would be alerted to its location using one of the three displays. The threats could appear in one of three sectors: either in front of the vehicle, to the side of the vehicle, or behind the vehicle. The participant was then required to orient the vehicle towards the location of the threat as quickly as possible using a joystick. Once the vehicle was oriented towards the threat, the participants would press the trigger. Localization error, the deviation of the vehicle heading from the location of the threat, and reaction time were measured for the primary task.

A secondary task was used to increase the difficulty of the task and to measure workload. Participants were required to listen for radio calls which were composed of colour and number combinations. Participants would acknowledge the call sign by pressing the corresponding button on a touch screen. Multiple radio calls could occur at the same time, and they were presented simultaneously with the threat appearances. Response time, defined by the time from trigger push (for the primary task) until radio call response, and proportion of correctly answered radio calls were computed for the secondary task.
A subjective questionnaire was answered by the participants at the end of the experiment.

Results

Participants' orientations to threats cued by the two displays with visual components (visual + audio display and the tri-modal display) had significantly lower localization errors than the tactile + audio display. The authors noted that the resolution of the localizations using the tactile belt was a lot lower than one would expect due to the 30 degree separations of the tactors. The authors suggested that this may be due to some temporal integration of the movement of the vibrating tactor as the participant re-oriented the vehicle. The authors also note that the HUD presentation did improve reaction time when compared to previous experiments, but the visual display needed to be in focal attentional to be useful while the tactile display could be used even if it was only in peripheral attention. The subjective ratings of the participants seemed to indicate that the 3D audio cue was not used as readily as the other cues, and the authors propose that this may be because of the auditory secondary task.

Conclusions:

A combination of multiple sensory presentations can overcome limitations of individual modalities. Audio radio secondary tasks are commonly used to increase task difficulty and workload, but it may interfere with presentation of auditory stimuli in the primary task.

Reference:

Tadema, J., & Theunissen, E. (2008). Design of a synthetic vision overlay for UAV autoland monitoring. *Proceedings of SPIE*, 6957, 69570B-69570B-11.

Overview:

This paper discusses the design and evaluation of a synthetic vision overlay for UAV autoland scenarios. Tadema and Theunissen state that in an UAV autoland scenario, operators are largely making use of rule-based behaviour (from Rasmussen's SRK taxonomy). This is because manual control of the vehicle is controlled by the autoland, which rules out skill-based control. Instead, the operator is responsible for monitoring flight variables and assessing if they fall within certain boundaries. Hence, Tadema and Theunissen hypothesize that the optimal role of a human operator is to "integrate and compare information from dissimilar sources." The authors propose that synthetic vision overlays are the best method for supporting this role for human operators.

Synthetic vision overlays superimpose projected flight paths onto the visual feed from a nose-mounted camera. This allows the operator to gauge how well the autoland system is operating by a simple visual comparison of the projected path and the visual stimuli. The synthetic vision overlays integrate trend information from both lateral and vertical tracking measures into a single visual element. The authors also considered different levels of automation control in the design of their interface (whether the automation can progress to the next stage without human intervention). Overall, there were two major design goals for the interface: the interface had to support conformance monitoring (making sure the autoland system was taking the correct steps to achieve the goal of landing), and integrity monitoring (making sure that there were no errors in the autoland system's calculations and sensors).

Methodology

An experiment involving monitoring of a simulated UAV autoland scenario was used to evaluate the effectiveness of the new synthetic vision enhanced interface. The independent variable was the interface type (new interface vs. conventional interface). The primary experimental task was to assess the integrity of the guidance information used by the autoland system during the approach. Participants could either allow the UAV to land or they could instruct the UAV to go-around. Both normal landing scenarios and abnormal positional data scenarios were used. In the abnormal scenarios, the autoland system would use incorrect geographical positioning information, which would cause it to land outside of the touch-down zone. No secondary tasks were used.

Results

The results showed that the new advanced interface reduced the variability in the go-around decisions made by the operators. The advanced interface was also able to increase the rate of correct identifications of integrity discrepancies without increasing the number of false alarms.

Conclusions: This paper described an experiment that has a very similar scenario as our project, involving an autoland scenario with a decision to abort or continue. While the data analysis was not explained in detail in this paper, it appears that hit rates and false alarm rates were calculated. The autoland scenario (with the abort/continue decision) is not a skill-based task, and relies largely on rule-based behaviour. The role of a human operator in the autoland situation is conformance monitoring and integrity monitoring.

Annex B Automation Problems in Commerical Aircraft

B.1 Uninhabited Aerial Vehicle Auto-landing Problems

Flight deck automation is the tasking of machines to perform operations that would have, until recently, been the role of the pilot. Some of these tasks include high workload (landing) as well as low workload (cruise flight) situations. Current flight deck automation includes autopilots, flight management systems, and warning and alerting systems. There are many parallels between automation in the commercial aircraft domain and the automation of the control of UAV control. As such, a review of problems that have affected automation in commercial aircraft has been provided.

The introduction of automation has seen a steady decline in aviation accidents. It has been generally well received by the pilot community and the transition from legacy or "round dial" machines to higher technology aircraft has been relatively uneventful. Nevertheless, with the advent of advanced technology, and the use of technology to monitor and actively react to safety critical functions, there are many who have expressed concern (Wiener, 1989) that technology has changed the role of the pilot from operator to monitor. Studies by Billings (1991; 1996) cited problems with flight deck automation and proposed a more human-centered approach to design and use. Sarter and Woods (1992; 1994; 1995) have sought to further investigate some of these pilot-led concerns into failure of pilot-automation interaction.

The fact that flight deck automation human factors issues exist is widely recognized. A comprehensive list of Human-Automation issues includes:

Technology breakdowns

- Automation may not work as desired under non-normal conditions
- Direct controls of automation may be poorly designed
- Displays may be poorly designed
- Failure modes may be unanticipated by designers
- Human-centered design philosophy may be lacking

Pilot reaction

- Automation behaviour may be unexpected or unexplained
- Automation may use different control strategies than pilots
- Mode awareness may be lacking
- Pilots may be out of the loop
- Pilots could become complacent
- Monitoring requirements may be excessive
- Information integration may be required

The initial assumption behind the use of automation was that it would reduce the quantitative nature of a task (Sarter & Woods, 1994). For example, for manned aircraft, automation would reduce the more mundane tasks of straight and level flight, or assume coordination of the aircraft for landing. This would allow the pilot to provide cognition actions toward some other task. In many cases, this task became monitoring of various aircraft systems. The resulting introduction of new technology had a surprising result. It did not reduce the overall task loading of the operator but did, however, change the nature of the loading. This dichotomy has been well documented in modern flightdeck design. The engineers designing flightdecks had little direct communication with operators so theory and practice were widely disjointed.

B.1.1 Accident Review

A review of accidents attributed to technology issues revealed some 16 accidents with varying loss of life (www.flightdeckautomation.com). The accidents consist of large, transport category aircraft and are collected from worldwide sources. A review of these accidents revealed common trends that could be extended to the use of Unmanned Aerial Vehicles (UAVs) and their operation. The problem of accidents involving high technology aircraft was highlighted by the crash of an Airbus A320 at Strasbourg, France.

The Airbus A-320 was the most highly automated civil aircraft flying at the time, and its introduction at the end of the 1980s was highlighted by the use of “fly-by-wire" control systems and an advanced Flight Management System (FMS). The fly-by-wire system was a first for transport category aircraft. In essence, control systems which were once directly connected from flight controls to control surfaces were replaced by computers. Electrical signals replaced pulleys and wires to affect deflection. The FMS is essentially a sophisticated autopilot capable of flying the aircraft along a pre-programmed path from takeoff to touchdown. It is essentially an “aircraft brain” that is the overseer of numerous bits of information to the aircraft state.

The ensuing investigation of the Strasbourg accident implied that the aircrew had inputted improper information into the FMS. Instead of commanding a descent rate of 3.2 degree angle in the "Flight Path Angle" mode, they had input a vertical descent rate of 3200-foot-per-minute in the “Vertical-descent” mode. The FMS did not interpret this as abnormal therefore no indication was given to the crew of anything out of the ordinary. The ensuing confusion over the state of the aircraft and subsequent crash became a famous incident as to what is now known as mode confusion (Hansman, 2001).

B.1.2 Mode Confusion

New technology is flexible in the sense that it provides practitioners with a large number of functions and options for carrying out a given task under different circumstances. However, this flexibility has a price. Because the human supervisor must select the mode best suited to a particular situation, their knowledge of the system operations must be more extensive than before. In addition, the human supervisors are also required to satisfy new monitoring and attentional demands to track which mode the automation is in and what the automation is doing to manage the underlying processes. When designers make use of multiple modes without supporting these new cognitive demands, it creates new mode-related errors and failure paths.

The accident of the A320, along with several others, revealed that pilots sometimes became confused about what the cockpit automation was doing. Consequently, an examination of these accidents was warranted. However, even a cursory look at the incident and accident data revealed more than just the inability of the crew to understand the automation. In *Aviation Automation: The Search for a Human-Centered Approach*, Charles Billings writes

> *"Today's flight management systems are "mode rich" and it is often difficult for pilots to keep track of them. The second problem, which is related to the first involves lack of understanding by pilot's of the system's internal architecture and logic, and therefore a lack of understanding of what the machine is doing, and why, and what it is going to do next."* (Billings, 1997)

The early stages of aviation automation were marked by a small number of independent modes. These modes may handle altitude, airspeed and heading. As technology advanced, the tendency was to incorporate "mode rich" systems that were interwoven. For example, modern transport aircraft are equipped with "path" modes that incorporate descent and speed constraints in tandem.

Another famous incident of mode confusion happened in April 26th 1994 with China Airlines flight 140 in Nagoya Japan. China Airlines Flight 140 was an Airbus A300 enroute from Taiwan to Nagoya. The flight was routine, however just before landing, the First Officer pressed the Takeoff/Go-around button (also known as a TO/GA) which increased thrust to a level of power that was required for take-off.

The co-pilot (who was flying the aircraft at the time) tried to correct the situation by manually controlling the thrust levers and forcing the control column down to reduce the climb rate. The autopilot, which thought it was in a takeoff scenario, responded to these actions by increasing the climb rate against the forces of the co-pilot. This nose-high attitude, combined with decreasing airspeed due to insufficient thrust, resulted in a stall of the aircraft. The subsequent crash killed 264 passengers and crew.

In their final report, Japan's Ministry of Transport cited a number of human factors engineering deficiencies that contributed to the crash. These included, "The captain and first officer did not sufficiently understand the FD (flight director) mode change and the AP (auto pilot) override function. It is considered that unclear descriptions of the AFS (Automatic Flight System) in the flight manual prepared by the aircraft manufacturer contributed to this." (JMOT, 1996) The first officer inadvertently triggering the TO/GA was another cause.

Automation confusion is only one part of the problem of supervisory control. If Control can be defined as to express "mastery" or "proficiency" of some skill or art (dictionary.com -2010), and proficiency assumes knowledge of that skill, then, by extension, that lack of knowledge then results in loss of control. This knowledge is common in many advanced technology aircraft accidents and is commonly referred to inert knowledge.

The NASA ASRS Database included a record of 276 incidents it classifies as “automation behaviour may be unexpected or unexplained”. This classification only provides a part answer. In examining accidents of this type, misinterpretation of aircraft “mode” could also be classified as “unexpected or unexplained”.

Another of the highest recorded incidents in the ASRS database is “understanding of automation may be inadequate”. This is confirmation that “Mode Confusion” has led to numerous incidents in Approach and Landing Scenarios.

B.1.3 Confirmation Bias

Confirmation bias is defined as “the tendency to prefer information that confirms their preconceptions or hypotheses, independently of whether they are true (Wikipedia, 2010). It is the theory that we create a solution that explains the situation and only seek out information that conforms to our understanding of it.

In aviation, several examples of confirmation bias have been recorded but not so famously as the case of British Midland flight 92. British Midland (BMI) flight 92 was a Boeing 737-400, on a scheduled flight from London Heathrow Airport to Belfast, Northern Ireland. Shortly after take-off from Heathrow, the left engine suddenly ruptured. The pilots, who were unaware as to the source of the problem, heard a loud noise and severe vibration, emanating from the back of the aircraft. In addition, smoke and burning fumes began pouring into the cabin via the ventilation system. Several passengers sitting near the rear of the plane noticed smoke and sparks coming from the left engine.

In consultation with the company and air traffic control, the flight was diverted to East Midlands airport. The captain, who was manually flying at this time, asked the First Officer which engine was malfunctioning, the First Officer replied: 'It's the le... it's the right one'. The engine gauges did not blatantly indicate which engine was malfunctioning.

In previous versions of the 737, the air conditioning ran through the right hand engine, but on the 737-400 it ran through both. The pilots had been used to the older version of the aircraft and did not realize that this aircraft, which was new to the airlines fleet, was different. The introduction of smoke into the cabin was an indication (on older B737s) that the smoke, and therefore failure, was coming from the right engine; this led them to shut down the working right engine instead of the malfunctioning left engine. They had no way of visually checking the engines from the cockpit, and the cabin crew did not inform them that smoke and flames had been seen from the left engine.

When the pilots shut down the right engine, they could no longer smell the smoke, which led them to believe that they had correctly dealt with the problem. This was an example of confirmation bias at work. As it turned out, this was simply a coincidence: when the autothrottle was disengaged to shut down the right engine, the fuel flow to the left engine was reduced and the excess fuel which had been igniting in the jet exhaust disappeared; therefore, the ongoing damage was reduced, the smoke smell ceased, and the vibration reduced, although it would still have been visible on cockpit instruments. The pilots, however, did not consult the vibration detectors because these instruments, on previous planes they had flown, were notoriously unreliable.

During the final approach to the East Midlands Airport, more fuel was pumped into the damaged engine to maintain speed, which caused it to cease operating entirely and burst into flames. The flight crew attempted to restart the right engine but it did not start in time.

B.1.4 Out-of-the-Loop (OOTP) Performance Degradation

The use of computers has changed the active role of pilots. By introducing this technology, the pilot's role has changed from active participant to manager of technology. As a consequence, the ability of pilots to understand and react to performance issues has been degraded. There is evidence, from both research and accident statistics, that people make poor monitors. For example, A laboratory study to compare failure detection performance found that the performance by participants who were actively controlling a dynamic system "was faster and more accurate" than the performance of those who were monitoring an autopilot that controlled the system. These results were attributed to the fact that in the manual mode, the participants remained in the "control loop" and benefited from the additional sensory cues derived from "hands on" interaction with the system. These findings agreed with a research study by L.R. Young (Funk, 1996). As well, system operators working with automation have been found to have a diminished ability both to detect system errors and subsequently to perform tasks manually in the face of automation failures, compared with operators who manually perform the same tasks. This "out-of-the-loop" degradation can be linked to two major issues

- loss of manual skills
- loss of situation awareness

B.1.5 Loss of Skill

The loss of manual skills is a major concern accompanying the introduction of automation. For Weiner and Curry (1980) found that supervisory controllers of automation were slower and more inefficient in bringing the system under control than were subjects who had operated only in a manual mode. They also expressed concerns by aircraft flight crews that a loss of proficiency will occur with extensive use of automatic equipment. The fear is that manual skills will degrade and that pilots will no longer be proficient at manual operations when needed.

This has become a concern in automated aircraft, where new pilots may have little opportunity to acquire or practice manual skills or may not take full advantage of the opportunities they do have (Orlady, 1989).

Evidence of the loss of simple flying skills can be seen in the following ASRS excerpt of an Airbus A320 crew that deviated from a precision approach into Atlanta Georgia.

“I began to descend on the glideslope. at this time, the line captain was looking up the tower freq. he looked up to notice my premature descent and heading overshoot and advised me to correct back to 4000 ft until on the loc. the tower controller, noting our pos, asked whether we had a problem and whether we had the airport in sight. The line captain replied that there was no problem but that there was training in progress and that we had the field in sight. The tower controller then cleared us for a visual approach to runway 26r. in the premature descent, I noticed a minimal altitude of 3650- 3700 ft msl, approx 15 degrees n of the localizer course at about 10-12 dme/atl. The approach was fully stabilized by 2000 ft msl (1000 ft agl). From that point, we completed an uneventful visual approach and landing on runway 26r. Human performance: I was well aware of the restriction and procedural requirements to maintain 4000 ft until established on the loc course. I allowed myself to become fixated on the glideslope indicator, subduing my awareness of localizer proximity that fixation also played a part in my overshooting the assigned intercept heading and detailed situational awareness”(ASRS.com-2010)

Several topics identified in the ASRS database (for example, programming errors, automation distraction, incorrect mode selection) account for 47% of incidents identified as automation issues. Although considered root causes of the various incidents, they all contributed to “loss of skill” manifestations.

B.1.6 Loss of Situation Awareness

The loss of situation awareness (SA) underlies a great deal of the out-of-the-loop performance problems (Endsley, 1987). Endsley defines situation awareness as "the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (Endsley, 1988, p. 97). From the definition of different SA states, (SA level 1,2,3), a loss of SA can be defined as those parameters that retard the cognitive process to project some future state based on the misdiagnosed cues of the present. This inability to recognize given information leads to the heart of the Automation issues in modern flightdeck design. This can be seen, as well as the degradation of flying skills, in the case of American Airlines 965.

American Airlines Flight 965, a Boeing 757, was a scheduled flight from Miami International Airport in Miami, Florida to Alfonso Bonilla Aragón International Airport in Cali, Colombia, which crashed into a mountain in Buga, Colombia on December 20, 1995, killing 151 passengers and 8 crew members. The crash was the first U.S.-owned 757 accident and the highest death toll of any accident in Colombia. It is also the highest death toll of any accident involving a Boeing 757 at that time. It was surpassed by Birgenair Flight 301 which crashed on 6 February, 1996 with 189 fatalities. It was the deadliest air disaster involving a U.S. carrier since the downing of Pan Am Flight 103 on December 21, 1988.

In the final report, the Columbian authorities concluded that the pilots of American Airlines 965 failed to use the automation correctly. When then became task saturated, they lost situational awareness and failed to execute a proper climb away from the mountains surrounding Cali. Specifically they concluded that:

- The use of the FMS was confusing and did not clarify the situation
- Neither pilot understood the steps necessary to execute the approach, even while trying to execute it
- Numerous cues were available that illustrated that the initial decision to accept runway 19 was ill advised and should be changed

Many of these issues deal with the behaviour of automation. Flightdeckautomation.com describes this behaviour as -- what they (the automation) are doing now and what they will do in the future based upon pilot input or other factors -- may not be apparent to pilots, possibly resulting in reduced pilot awareness of automation behaviour and goals.

With respect to the accident in Cali, investigators recommended the need for automation to confirm changes manually made by the pilots. The hope here is that a high level of Situational Awareness is maintained throughout the flight. This was a contributory factor to the accident in Columbia. Upon changing the Beacon identifier, a significant course change resulted in the aircraft being flown into the side of the mountains surrounding the city.

B.1.7 Conclusion

Automation induced accidents in the landing regime have been attributed to a wide variety of root causes. Examination of these incidents reveals many common factors that exist across cultural and workplace environmental lines.

Given the breadth of latent causes, the challenge for systems designers is to try and identify how operators recognize the situations and react in a correct and timely manner.

This page intentionally left blank.

List of symbols/abbreviations/acronyms/initialisms

AEC	Audiological Engineering Corporation
AH	Abstraction Hierarchy
ANOVA	Analysis of Variance
APU	Auxiliary Power Unit
BD	Burst Duration
CF	Canadian Forces
COA	Course of Action
CTA	Control Task Analysis
CTS	Critical Task Sequences
CWA	Cognitive Work Analysis
dBSL	Decibels above Sensation Level
DND	Department of National Defence
DRDC	Defence Research & Development Canada
EAI	Engineering Acoustics Inc
EEG	Electroencephalography
EID	Ecological Interface Design
ERP	Event-Related Potentials
fMRI	Functional Magnetic Resonance Imaging
GCS	Ground Control Station
HGA	Hierarchical Goal Analysis
HH	High-Spatial & High-Temporal
HL	High-Spatial & Low-Temporal
IAI	Intelligent Adaptive Interfaces
IBI	Inter Burst Interval
ISR	Intelligence, Surveillance and Reconnaissance
ISTAR	Intelligence, Surveillance, Target Requisition, and Reconnaissance
IVIS	In-Vehicle Information Systems
KBB	Knowledge-Based Behaviour
LH	Low-Spatial & High-Temporal
MATB	Multi-Attribute Task Battery

MAUVE	Multiple Autonomous Uninhibited Vehicle Experimental
MMWS	Multimodal Workstation
MRT	Multiple Resource Theory
M-SWAP	Multisensory Assessment Protocol
OMI	Operator Machine Interface
Pre-SD phase	Pre-Spatial Disorientation phase
R&D	Research & Development
RBB	Rule-Based Behaviour
RMS	Root Mean Square
StA	Strategies Analysis
SA	Situation Awareness
SAGAT	Situation Awareness Global Assessment Technique
SBB	Skill-Based Behaviour
SOA	Stimulus Onset Asynchrony
SOP	Standard Operating Procedures
SOW	Statement of Work
SRK	Skills, Rules, Knowledge
TCCTA	Temporal Coordination Control Task Analysis
TLS	Tactor Locator System
TOGA	Turn-Off/Go-Around
TSAS	Tactile Situation Awareness System
TTC	Time-to-Collision
UAS	Uninhabited Aerial Systems
UAV	Uninhabited Aerial Vehicles
WDA	Work Domain Analysis
FMS	Flight Management System
AP	Auto-pilot
AFS	Automatic Flight System
FD	Flight Director
BMI	British Midland
CRM	Coordinate Response Measure
DND	Department of National Defence

DRDC	Defence Research & Development Canada
DRDKIM	Director Research and Development Knowledge and Information Management
R&D	Research & Development

This page intentionally left blank.

Glossary

Agent Model: This model incorporates knowledge relating to the participants of the system (i.e., computer and human agents), as well as their roles and responsibilities

Analogous Icon: An icon that visually captures a constraint in the environment.

Attentional Mapping: Step that is important when designing modalities that cannot be ignored, and that have strong temporal qualities.

Audification: Straight signal-to-sound conversion; translation of some physical stimuli into an auditory representation.

Auditory Icon: Sounds which have a direct link to a real world object or event (such as footsteps).

Auditory Signals: (Earcons, auditory icons, audifications, and sonification) provide a ripe lexicon of perceptual signals that can be used by designers to support SBB, RBB, and KBB.

Backward masking: When the target stimulus is corrupted with a subsequently presented masking stimulus.

Burst Duration (BD): Which is the time between the onset and end of a burst.

Cognitive Work Analysis (CWA): A constraints based framework for analyzing complex systems.

Control Task Analysis (CTA): The second phase in cognitive work analysis. Describes and models how a task is accomplished.

Data Visualization: An image constructed to convey information about data

Decibels above sensation level (dBSL): Measures the amplitude of a signal relative to an individual's sensation threshold.

Design Model: This model comprises the hardware and software requirements related to the construction of the intelligent adaptive system. This model also specifies the means by which operator state is monitored.

Dialogue/Communication Model: This model incorporates knowledge of the manner in which communication takes place between the human operator and the system, and between the system agents themselves.

Earcon: Sounds which do not have a direct link to the real world but can be arranged to communicate information.

Ecological interface design: Design approach that has been used to great success in complex socio-technical systems.

Ecological valid tactile patterns: Tactile stimuli that produces an easily recognizable real-world sensation. Not a formal term, and has not be explored in detail within the literature.

Endogenous attention: Refers to the voluntary control of attention. Governed by goal-driven attentional control, which is associated with the response to symbolic cues. These symbolic cues are associated with stimuli that indirectly point to a potential target location.

Exogenous attention: Refers to attention being drawn without conscious attention. Governed by stimulus-driven attentional control, this is associated with the response to perceptual characteristics of the stimuli instead of the semantic meaning of the stimuli.

Forward masking: When the target stimulus is corrupted with a preceding masking stimulus.

Icons: Graphic symbols that represent a concept or process due to the similarities between the graphical element and its real-world equivalent.

Intelligent Adaptive Interfaces (IAI): A system that adjusts the machine's characteristics and/or display to dynamically change with external events in terms of operator states and mission goals in real time.

Knowledge-Based Behaviour (KBB): Represent the work domain in the form of an abstraction hierarchy to serve as an externalized mental model that will support knowledge-based problem solving.

Knowledge Model: This model incorporates a detailed record of the knowledge required to perform the tasks that the system will be performing.

Load Stress: Stress caused by increasing the number of channels over which is information is presented.

Mechanoreceptors: Type of stimuli which are sensitive to pressure, vibration, and slip.

Meissner Corpuscles: A stack of nerve fibres, located in the grooved projections of the skin surface formed by epidermal ridges, situated perpendicular to the skin surface. They respond to light touch and are velocity sensitive. They are sensitive to vibrotactile stimuli in the range of 10 – 100Hz. They have highest sensitivity (lowest threshold) when sensing vibrations less than 50Hz. Meissner corpuscles are categorized as rapid adapting (RA) receptors which respond quickly to a stimulus, but rapidly adapt to it and stop responding when subjected to a constant stimulus.

Merkel Receptors: Disk shaped receptors that respond to pressure and texture, but also to low frequency (5-15 Hz) vibratory input. They are categorized as slow adapting (SA) receptors which adapt slowly to stimulus and continue to transmit when subjected to constant pressure. Tactile display systems, by necessity, are in constant contact with the skin and are not well suited for the stimulation of SA type receptors.

Multiple Resource Theory (MRT): Encompasses the independent modality-specific attentional resource theory. The main premise of MRT is that humans do not have a single source capable of information processing, but a number of resources that can be accessed concurrently.

Nocioreceptors: Type of stimuli which are pain receptors.

Organization Model: This model incorporates knowledge relating to the organizational context that the knowledge-based system is intended to operate in (e.g. command and control (C2) structures, Intelligence Surveillance, Target Requisition and Reconnaissance - ISTAR etc.)

Pacinian Corpuscles: The largest receptors of the skin. These are located deeper in the skin and most susceptible to the vibrations in the 200-350 Hz frequency range. Pacinian corpuscles are categorized as RA receptors. This means that the effect of stimuli degrades rapidly after onset. Pacinian corpuscles discharge only once per stimulus application, hence they are not sensitive to steady pressure.

Peripersonal space: The space immediately surrounding the body; the space where objects can be grasped and manipulated.

Proprioceptors: Type of stimuli which give information about the position of the limb in space.

Ruffini Corpuscles: Spindle shaped receptors that respond to skin stretch and mechanical deformation within joints, specifically angle changes up to 2 degrees. They contribute to providing feedback for the grip and grasping function. These are categorized as SA receptors and are located in the deep layers of the skin.

Rule-Based Behaviour (RBB): Provide a consistent one-to-one mapping between the work domain constraints and the cues or signs provided by the interface.

Semantic mapping: Process where variables are mapped into perceptual characteristics. This process is fundamental to fulfilling the 2nd EID principle where constraints should be mapped onto perceptual objects.

Signal visualization: A translation of some physical stimuli into a visual representation.

Sonification: Mapping of information to sound parameters to create the auditory equivalent of visualization.

Skill-Based Behaviour (SBB): To support interaction via time-space signals, the operator should be able to act directly on the display and, the structure of the displayed information should be isomorphic to the part-whole structure of movements.

Skills, rules, knowledge (SRK) taxonomy: Describes different levels of cognitive control. Operators of complex systems are capable of using control strategies based on Skill-Based Behaviour (SBB), Rule-Based Behaviour (RBB), or Knowledge-Based Behaviour (KBB).

Spatio-temporal tactile patterns: A pattern created by the sequential activation of a series of vibrotactors to intuitively present information using multiple dimensions.

Speed Stress: Stress caused by changing the rate of signal presentation.

Steven's Power Law: The relationship between changes in an objective parameter of an auditory alarm (e.g. pitch or speed) and the subjective perception of the urgency.

Stimulus Onset Asynchrony (SOA): The time between the onsets of two consecutive bursts.

Symbolic-analogic continuum: Describe symbolic displays as ones that establishes mapping between a sound and an intended meaning, with no intrinsic relationship existing.

System Model: This model incorporates knowledge of the system's abilities, needs, and the means by which it can assist the human operator (e.g., advice, automation, interface adaptation).

Tactification: A translation of some physical stimuli into a vibro-tactile representation. This is not a formal term, and has not been studied in detail in the literature.

Tactons: Brief messages that can be used to represent complex concepts and information in vibrotactile displays. They are categorized in three main groups; compound *tactons*, hierarchical *tactons* and transformational *tactons*.

Task Model: This model incorporates knowledge relating to the tasks and functions undertaken by all agents, including the operator.

Temporal masking: When the vibrations are presented to the same location, and the target stimulus is presented either within the time interval of the masking stimulus, or near the onset or just after the offset of the masking stimulus.

Two-point Discrimination: Minimum distance between two stimuli to be perceived as two distinct stimuli instead of one large stimulus.

Thermoreceptors: Type of stimuli which are sensitive to changes in temperature.

User Model: This model incorporates knowledge of the human operator's abilities, needs and preferences.

Visual Thesaurus: Set of visual forms that can be used to represent work domain properties. The visual forms used include visual primitives (bar graphs and other simple iconic elements), complex combinations of visual primitives (connections, grouping, etc.).

Weber Fraction: A formula that is often used to determine the minimum threshold of perceived change in any parameter (e.g., amplitude, frequency, weight). For frequency, it is the differential threshold divided by the reference frequency, expressed as a percentage.

Work Domain Analysis (WDA): An analysis stage used in CWA and EID to capture the physical and functional properties of a work domain.

World Model: This model incorporates knowledge of the external world, such as physical (e.g. principles of flight controls), psychological (e.g., principles of human behaviour under stress), or cultural (e.g., rules associated with tactics adopted by hostile forces).

DOCUMENT CONTROL DATA
(Security classification of the title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (The name and address of the organization preparing the document, Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's document, or tasking agency, are entered in section 8.) Publishing: DRDC Toronto Performing: University of Waterloo Advanced Interface Design Lab,200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1 Monitoring: Contracting:	2. SECURITY CLASSIFICATION (Overall security classification of the document including special warning terms if applicable.) UNCLASSIFIED

3. TITLE (The complete document title as indicated on the title page. Its classification is indicated by the appropriate abbreviation (S, C, R, or U) in parenthesis at the end of the title)

Multimodal Interfaces: Literature Review of Ecological Interface Design, Multimodal Perception and Attention, and Intelligent Adaptive Multimodal Interfaces (U)
(U)

4. AUTHORS (First name, middle initial and last name. If military, show rank, e.g. Maj. John E. Doe.)

Wayne Giang; Sathya Santhakumaran; Ehsan Masnavi; Doug Glussich; Julianne Kline; Fiona Chui; Catherine Burns; Jonathan Histon; John Zelek

5. DATE OF PUBLICATION (Month and year of publication of document.) August 2010	6a NO. OF PAGES (Total containing information, including Annexes, Appendices, etc.) 269	6b. NO. OF REFS (Total cited in document.)

7. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of document, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Contract Report

8. SPONSORING ACTIVITY (The names of the department project office or laboratory sponsoring the research and development − include address.)

Sponsoring:
Tasking:

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant under which the document was written. Please specify whether project or grant.)	9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.) w7711−098148/001/TOR
10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document) DRDC Toronto CR 2010−051	10b. OTHER DOCUMENT NO(s). (Any other numbers under which may be assigned this document either by the originator or by the sponsor.)

11. DOCUMENT AVAILABILITY (Any limitations on the dissemination of the document, other than those imposed by security classification.)

Unlimited distribution

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11), However, when further distribution (beyond the audience specified in (11) is possible, a wider announcement audience may be selected.))

Unlimited announcement

DOCUMENT CONTROL DATA

(Security classification of the title, body of abstract and indexing annotation must be entered when the overall document is classified)

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

(U) To improve operational effectiveness for the Canadian Forces (CF), the Joint Unmanned Aerial Vehicle Surveillance Target Acquisition System (JUSTAS) project is acquiring a medium−altitude, long−endurance (MALE) uninhabited aerial vehicle (UAV). In support of the JUSTAS project, Defence Research and Development Canada (DRDC) – Toronto is investigating the human factors issues of UAV ground control stations (GCS) interfaces for UAVs and exploring possible solutions using multimodal displays. This report analyzes current literature on multimodal perception and psychology in the context of developing a GCS simulator to evaluate the efficacy of multimodal displays for controlling UAVs. The report discusses the application of Ecological Interface Design (EID) to multimodal interface development, multimodal information presentation in non−visual modalities, and issues and implications of using multiple sensory modalities (e.g. crossmodal effects). In addition, the role of Intelligent Adaptive Interfaces (IAI) with respect to multimodal interfaces and current problems with automation in commercial aircraft are addressed. Recommendations are provided to develop a program of research to enhance the design of GCS interfaces to support future requirements of the JUSTAS project.

(U) En vue d'améliorer l'efficacité opérationnelle des Forces canadiennes (FC), l’acquisition d’un engin télépiloté (UAV) moyenne altitude et longue endurance (MALE) est un des volets du projet Système interarmées de surveillance et d'acquisition d'objectifs au moyen de véhicules aériens sans pilote (JUSTAS). À l’appui du projet JUSTAS, Recherche et développement pour la défense Canada (RDDC) — Toronto effectue des recherches sur les problèmes relatifs aux facteurs humains des interfaces UAV pour les postes de contrôle au sol (PCS) d'UAV et sur les solutions possibles au moyen d’affichages multimodaux. Le présent rapport porte sur l'analyse de littérature existante sur la perception et la psychologie multimodales dans le cadre du développement d'un simulateur PCS en vue d'évaluer l'efficacité d'affichages multimodaux pour commander les UAV. Le rapport comporte également un examen de l'application de la conception d'interfaces écologiques (EID) au développement d'interfaces multimodales, de la présentation d'information multimodale dans des modes non visuels et de problèmes et répercussions relatifs à l'utilisation de modes sensoriels multiples (p. ex. effets intermodaux). Le rôle d'interfaces adaptatives intelligentes par rapport aux interfaces multimodales et les problèmes actuels avec l'automatisation à bord des aéronefs commerciaux sont également abordés. De plus, des suggestions relatives à la mise au point d'un programme de recherches visant à améliorer la conception des interfaces PCS à l’appui des exigences futures du projet JUSTAS sont faites.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

(U) uninhabited aerial vehicle; multimodal display; ground control station interface

Defence R&D Canada

Canada's Leader in Defence
and National Security
Science and Technology

R & D pour la défense Canada

Chef de file au Canada en matière
de science et de technologie pour
la défense et la sécurité nationale

DEFENCE R&D DÉFENSE

www.drdc-rddc.gc.ca